The Bell System Technical Journal

January, 1932

The Cathode Ray Oscillograph *

By J. B. JOHNSON

The cathode ray oscillograph, since its invention by Braun, has developed along three lines. The major types of tubes are the high voltage tubes with a fluorescent screen, the high voltage tubes with internal photographic equipment, and the low voltage tubes. This paper follows the structural development of commercial tubes. The operation of the tubes is discussed, from the standpoint of both theory and practice, with particular reference to the low voltage type of tube. Numerous examples are given of the applications of the tubes to problems in science and engineering.

IN our complicated life, we find that we need a great many aids to our primary sense organs. The processes of the modern world demand that we make correct estimates of things that are too large or too small, too intense or too feeble, for our poor senses. We have balances to give us the weight of masses too heavy for us to lift or too small to be felt. Telescopes enable us to see far-off objects, microscopes very small objects. Our ears are supplemented by telephones that put us within earshot of almost all the civilized world. For electric currents we have ammeters to measure currents so large as to destroy us in a second, and galvanometers that measure currents far too small for us to feel as a shock. Taste and smell have not yet been supplied with artificial aids, but that may come some day.

For recording long times we have clocks and calendars; for making a record of happenings that take place in a time too short for us to think of, we use oscillographs.

There are a number of different types of oscillographs in use, all of them electrical in nature. The kind I am going to discuss involves a stream of cathode rays and it is therefore called the cathode ray oscillograph. The principle of its operation is quite simple. We have two electrodes in an elongated, evacuated glass tube as in Fig. 1; one of them may be a heated filament, the other a plate with a small hole in it. When a potential is applied between the electrodes, making the filament cathode and the plate anode, the electrons emitted by the hot filament are drawn to the anode. Some of them pass through the fine hole in the anode and continue as a thin pencil of electrons, a cathode

* Presented at Franklin Inst. mtg., Dec. 4, 1930. *Jour. Franklin Inst.*, December, 1931.

ray, down the length of the tube. At the end of the tube is a screen of fluorescent materials, which shines brightly at the point where the ray strikes it. We can therefore see where the ray ends on the screen. Another pair of electrodes in the form of two plates P and P_1 is introduced so that the cathode ray passes between them (Fig. 2). Now,

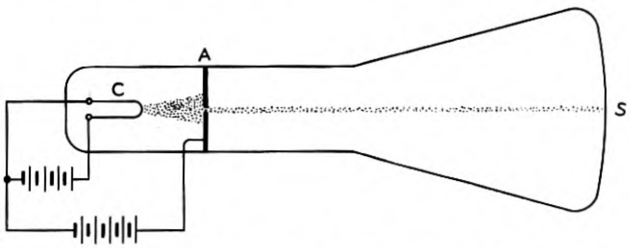


Fig. 1—Schematic of cathode ray tube.

by a battery or otherwise, we apply a voltage between the plates, so that one is positive with respect to the other. The electrons of the ray, being negative charges, are during their passage between the plates drawn toward the positive plate and emerge in a different direction because of the applied voltage. Similarly, a magnetic field applied by the magnet N-S, across the path of the ray and in the plane of the paper, as in Fig. 3, makes the ray emerge in a direction out from the

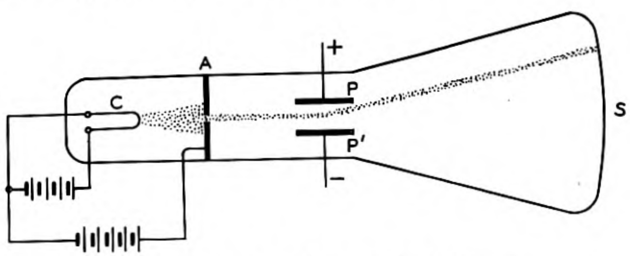


Fig. 2—Tube with electrostatic deflection.

page. The amount of the deflection is a measure of the strength of the applied magnetic or electric field. We have, then, in this cathode ray a pointer which tells the magnitude of the field that deflects it. It is, furthermore, a pointer that is almost without mass and sluggishness; it is almost not a material pointer. It can therefore follow variations in the applied field that are very rapid, as we shall see presently. Because of this property the instrument has been used extensively for studying the electrical phenomena of such a range of subjects as power machinery, telephone apparatus, radio transmission and electric surges.

One of the more dramatic recent applications has been the investigation of lightning, probably the most important work on lightning since its electrical nature was discovered by Benjamin Franklin one hundred eighty years ago.

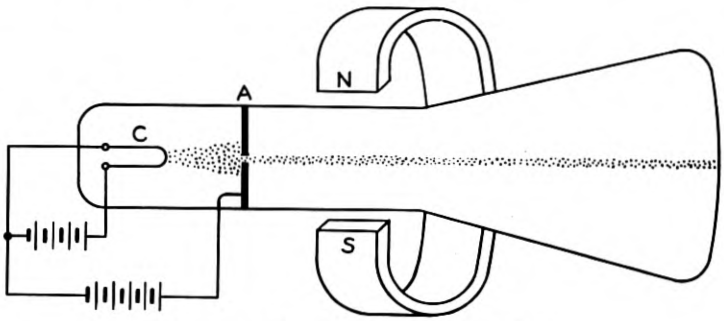


Fig. 3—Tube with magnetic deflection.

Let us examine the operation of the tube more closely. The speed of the electrons as they emerge from the aperture in the anode can be determined from the energy equation,

$$\tfrac{1}{2}mv^2 = eV,$$

the energy of motion equaling the total work done on the electron by the field between cathode and anode. V is the potential between cathode and anode, e the electric charge constituting the electron, m its mass and v its speed. Solving for the speed we have the relation between the speed and the driving voltage,

$$v = \sqrt{2\frac{e}{m}V}.$$

The value of e/m is known to be 1.77×10^7 e.m.u., the volt is 10^8 e.m.u. and the velocity of the electrons is thus

$$v = 5.95 \times 10^7\sqrt{V}\text{ cm./sec.}$$

If the driving potential is 300 volts, then the speed of the electrons is given as roughly 1×10^9 cm./sec. or 6000 miles per second. For a tube 20 cm. long an electron travels from the deflector plates to the screen in $20/10^9 = 1/50{,}000{,}000$ sec. If the applied voltage is 30,000 volts the speed is very nearly ten times as great as with 300 volts; it is ⅓ the velocity of light. A change of direction of the ray induced at the deflector plates is therefore transmitted to the end of the ray in a very short time, and the ray can follow faithfully potential variations at the plates that are very rapid.

Let us see how the ray responds to voltage applied to the plates. The ray normally travels with a speed v along the tube. Referring to Fig. 4, the ray now passes between two plates of length l and sepa-

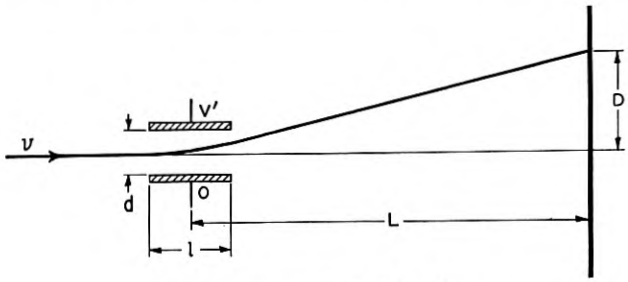


Fig. 4—Electrostatic deflection.

ration d, between which a potential difference V' is maintained. While the ray is passing between the plates the electrons are subject to an acceleration,

$$a = \frac{e}{m} E = \frac{e}{m} \frac{V'}{d}.$$

This continues during the time $t = l/v$. The transverse velocity acquired is therefore

$$v' = at = \frac{e}{m} \frac{V'}{d} \frac{l}{v}.$$

The ray then travels on in a straight line to the screen which it meets at a distance D from the normal position. The deflection D bears the same relation to the length of the beam, from the center of the deflecting plates, as the transverse velocity bears to the longitudinal.

$$\frac{D}{L} = \frac{v'}{v} = \frac{e}{m} \frac{V'}{d} \frac{l}{2 \frac{e}{m} V} = \tfrac{1}{2} \frac{l}{d} \frac{V'}{V},$$

$$D = \tfrac{1}{2} \frac{lL}{d} \frac{V'}{V}.$$

This brings out an interesting point in connection with the design of the deflector plates. For high sensitivity the plates should be long and close together but the plates must not cut the path of the deflected ray. If we want to get a certain maximum deflection D with a tube of a certain length L, then the relation of spacing to length of the plates must be

$$\frac{d}{l} = \frac{D}{L},$$

as can easily be seen from Fig. 4. When this condition is satisfied it makes little difference in the sensitivity whether the plates are large and far apart or small and close together.

The magnetic sensitivity of the tube is more uncertain only because the boundaries of the magnetic field are usually less well defined than those of the electric field. One form of the derivation will be given here. If the ray generated by the driving voltage V passes for a distance l (Fig. 5) through a region in which there is a transverse magnetic field of strength H, the path of the beam is an arc of a circle whose radius is

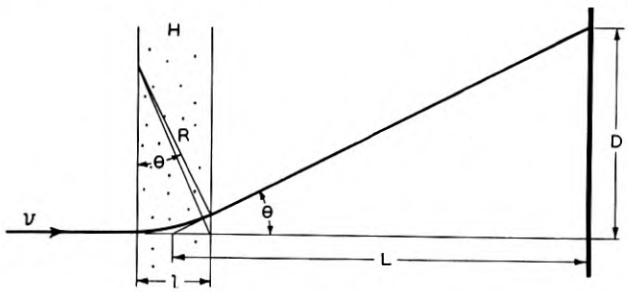


Fig. 5—Magnetic deflection.

$$R = \frac{mv}{eH} = \frac{1}{H}\sqrt{\frac{2mV}{e}}.$$

When the ray leaves the magnetic region it proceeds in a straight line to the screen, where the total deflection from the normal is D. If the angular deflection θ is not great then we have, very nearly,

$$\tan\theta = \frac{D}{L} \doteq \frac{l}{R} = lH\sqrt{\frac{e}{2mV}}$$

and hence

$$D \doteq lLH\sqrt{\frac{e}{2mV}}.$$

In practical units instead of electromagnetic the expression becomes

$$D \doteq .3lLH/\sqrt{V}.$$

Having now described in an elementary way how the cathode ray oscillograph works, let me turn to the story of the development of the tube.[1] The first reference I have seen to the idea that a cathode ray might be used to indicate magnetic field dates to 1894, when Hess,[2] in France, suggested the use of such a tube as a curve tracer. The first application of the idea, however, was made by Ferdinand Braun [3] in 1897, and after him the instruments have been called Braun tubes.

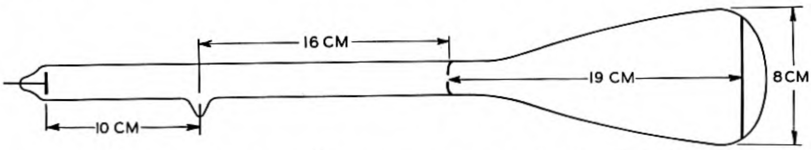


Fig. 6—The first cathode ray oscillograph, F. Braun, 1897.

The tube of Braun was quite simple (Fig. 6). It had a flat disc cathode, a wire in a side tube as anode, a pierced diaphragm to limit the beam and a fluorescent screen of zinc sulphide. It contained air at low pressure. Current from an electrostatic machine produced a discharge in the residual gas in the tube, from which emanated the cathode rays through the aperture in the diaphragm.

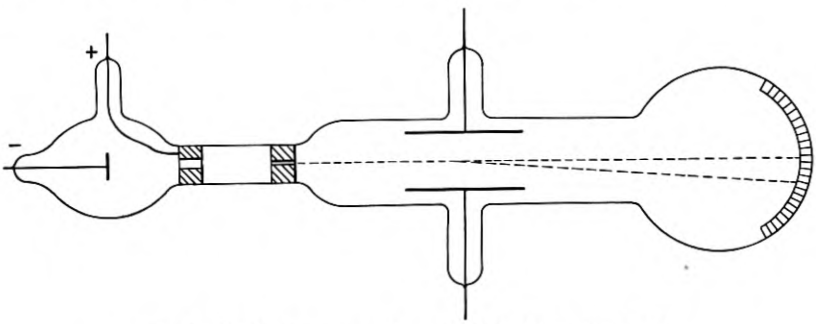


Fig. 7—Tube for measuring e/m, J. J. Thomson, 1897.

It is of interest that the invention of the tube took place before the nature of cathode rays was understood. It was in the same year that J. J. Thomson in England and W. Kauffmann in Germany, each using a tube that was almost identical with the Braun tube, deter-

[1] More detailed chronicles of the development of the tube have been made by H. Hausrath, "Apparatus and Technique for Producing and Recording Curves of Alternating Currents and Electrical Oscillations," *Helios*, 1912; and by MacGregor-Morris and Mines, "Measurements in Electrical Engineering by Means of Cathode Rays," *Jl. Inst. El. Eng.*, **63**, p. 1056, 1925.
[2] Hess, A., *Compt. Rend.*, **119**, p. 57, 1894.
[3] Braun, Ferdinand, *Wied. Ann.*, **60**, p. 552, 1897.

mined that cathode rays have mass as well as charge. Thomson's tube [4] is shown in Fig. 7.

The Braun tube immediately found many applications. One of the most fruitful fields for it was its use by Professor Zenneck in studying radio circuits and the transmission of radio waves. From Professor Zenneck and his school there are still papers coming out on work done by means of the Braun tube. The tube was introduced in this country

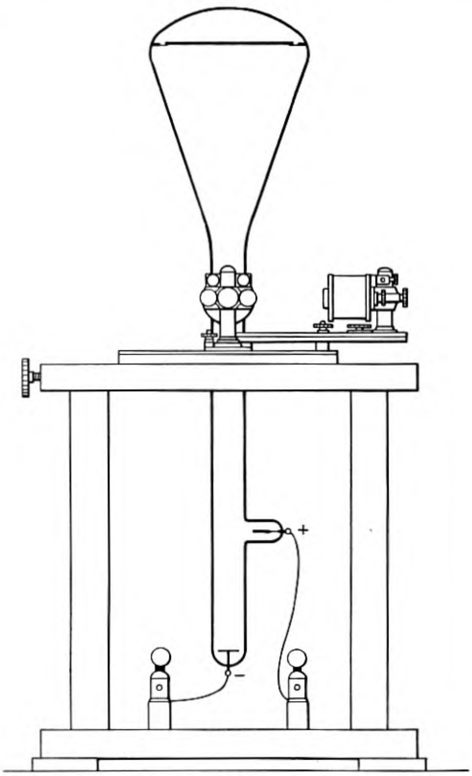


Fig. 8—Ebert and Hoffmann, 1898—tube by Geissler.[5]

very early. Professor H. J. Ryan, then at Cornell, in 1903 described measurements on high voltage power circuits, and similar work has appeared occasionally ever since from Professor Ryan's hand.

After the invention of the tube there were a number of improvements that made it more convenient and its operation more reliable. Figures 8–12 show some of the many designs of tubes of this time.

[4] Thomson, J. J., *Phil. Mag.*, **44**, p. 293, 1897.
[5] Ebert and Hoffmann, *E. T. Z.*, **19**, p. 405, 1898.

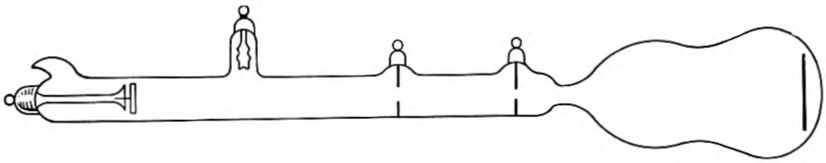


Fig. 9—MacGregor-Morris, 1902—tube by Cossor.[6]

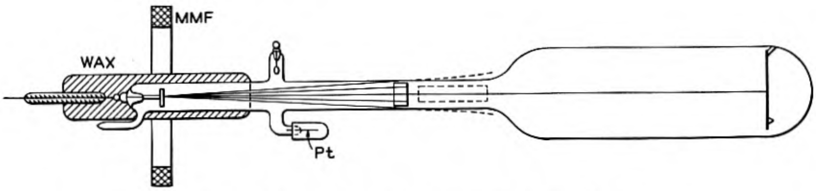


Fig. 10—Ryan, 1903—tube by Muller-Uri.[7]

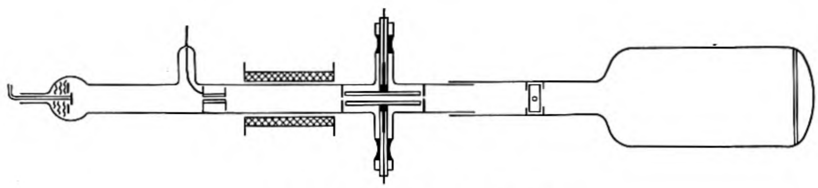


Fig. 11—Roschansky, 1911.[8]

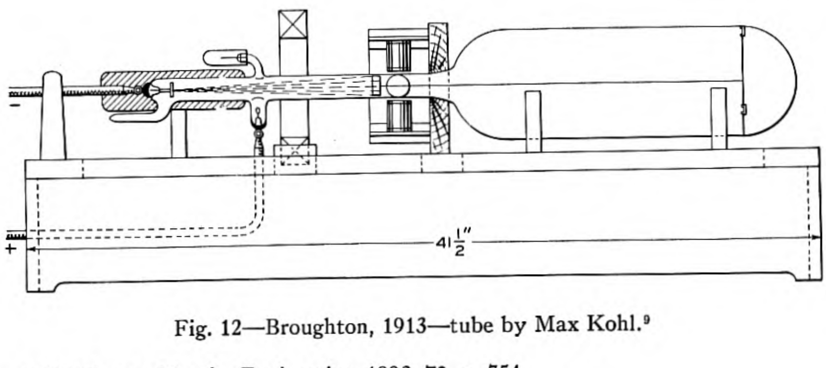


Fig. 12—Broughton, 1913—tube by Max Kohl.[9]

[6] MacGregor-Morris, *Engineering*, 1902, **73**, p. 754.
[7] Ryan, H. J., *Am. Inst. El. Eng., Trans.*, **22**, p. 539, 1903.
[8] Roschansky, D., *Ann. d. Phys.*, **36**, p. 281, 1911.
[9] Broughton, H. H., *Electrician*, **72**, p. 171, 1913.

In 1905 Wehnelt [10] suggested the use of a hot, lime coated filament, which he had found a couple of years earlier to be a strong emitter of electrons and which was the basis of the present day oxide coated filaments in vacuum tubes. Wehnelt made up such a tube (Fig. 13) which he could operate on the 220 volt power circuit. This was probably the first practical application of the oxide coated filament. A number of experimental tubes were made up with hot filaments in the

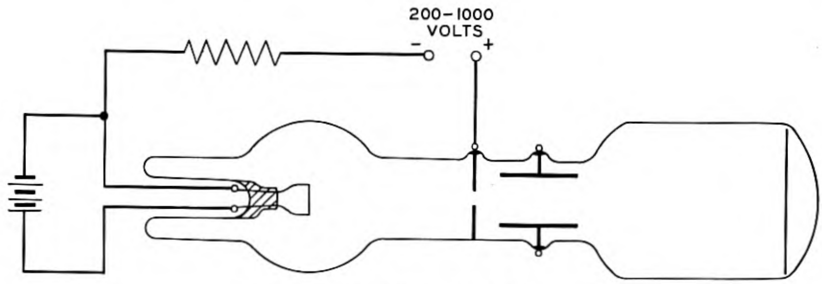


Fig. 13—Wehnelt, 1905.

years following, but almost twenty years were to elapse before a really successful tube with hot cathode and low voltage was developed. This was the Western Electric 224 tube,[11] to be discussed in greater detail presently.

Another tube of this type, by von Ardenne and Hartel,[12] is illustrated in Fig. 14.

After Braun and Wehnelt, the most notable change in structure was made by A. Dufour [13] in France in 1914. Up to that time permanent records of the pattern on the fluorescent screen were made by means of a camera. This usually meant that the pattern for a rapid phenomenon had to be repeated many times before the photographic plate was exposed enough. Dufour omitted the fluorescent screen and instead placed the photographic plate inside the tube so that the cathode ray could play directly on it. When the cathode rays strike the photographic emulsion directly a record can be traced in a much shorter time than when the intermediary light from a fluorescent screen is focused by a lens on the photographic plate. Placing the photographic plates internally of course involved a number of complications, such as mechanism for moving the plates inside the evacuated tube, means for inserting and taking out the plates, and pumps for producing and maintaining the vacuum. The old glass structure

[10] Wehnelt, A., *Phys. Zeit.*, **6**, p. 732, 1905.
[11] Johnson, J. B., *Jl. Am. Opt. Soc. & R. S. I.*, **6**, p. 701, 1922.
[12] Hartel, H. von, *Zeit. f. Hochfr. Techn.*, **34**, p. 227, 1929.
[13] Dufour, A., *Compte Rend.*, **158**, p. 1339, 1914.

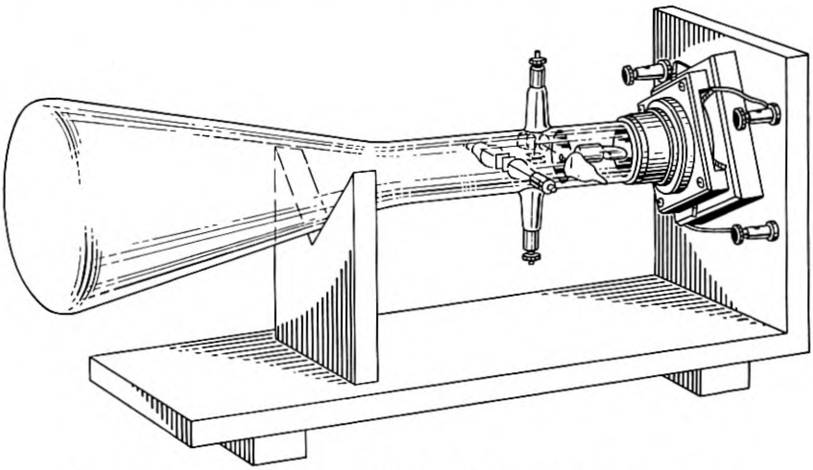


Fig. 14—von Ardenne-Hartel, 1930—tube by Leybold's Nachfolger.

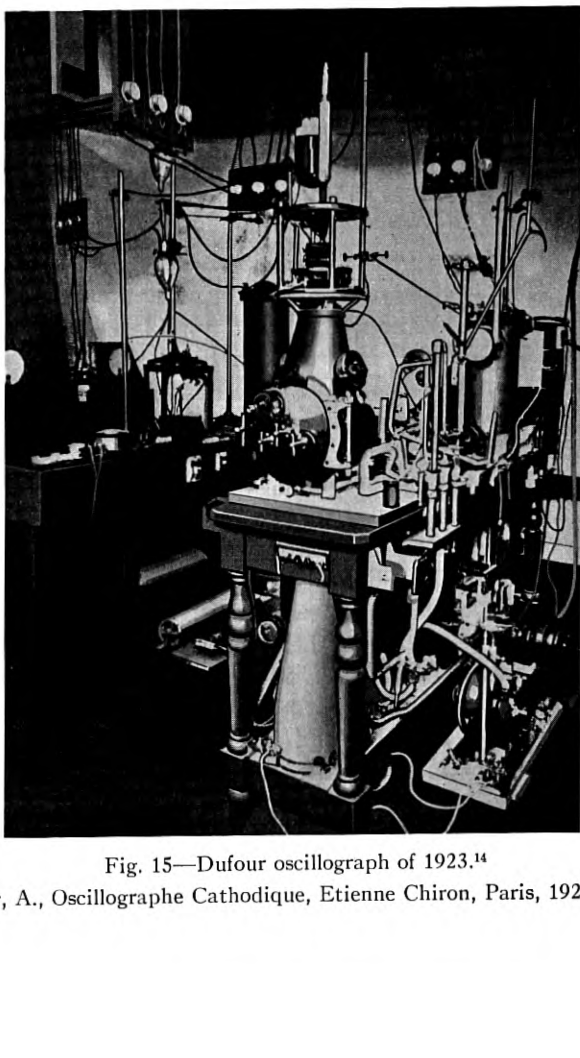


Fig. 15—Dufour oscillograph of 1923.[14]

[14] Dufour, A., Oscillographe Cathodique, Etienne Chiron, Paris, 1923.

is largely abandoned in these tubes, and metal is substituted. The applied voltage is very high, of the order of 50,000 volts. This makes a rather formidable piece of apparatus (Fig. 15), but quite a useful one.

In the last few years several different tubes of this type have been developed. Aside from Dufour's tube, there are those of Rogowski in Germany, Wood in England, Berger in Switzerland, Norinder in Sweden, The Westinghouse Co. and the General Electric Co. in this

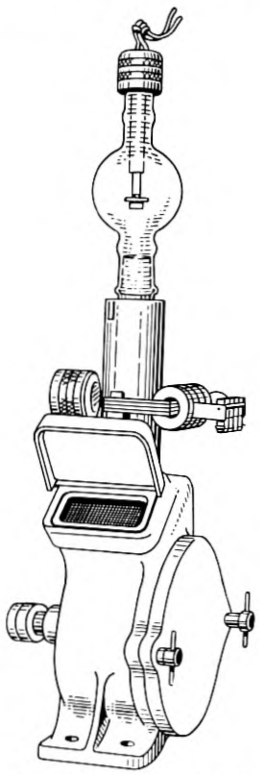


Fig. 16—Wood, 1923.[15]

country (Figs. 16–18, 21). All of them involve complicated tubes and control circuits. Some tubes are made to operate during a single peak of a 60 cycle wave. With others there are ingenious switching devices that start the tube operating at the very beginning of the electrical impulse to be studied, and the tube then proceeds to record the rest of the impulse. When we consider that the impulse may be a stroke of lightning on a transmission line, we realize that there are some things that are "faster than greased lightning."

[15] Wood, A. B., *Phys. Soc. Lond. Proc.*, **35–2**, p. 109, 1923.

In the last couple of years a further step has been made by Max Knoll.[16] He seems to have simplified the operation of the tube considerably by attaching on the end of his tube a thin window in the manner of Dr. Coolidge's cathode ray tube. In this way the photographic plate on the outside of the tube is exposed to the pencil of

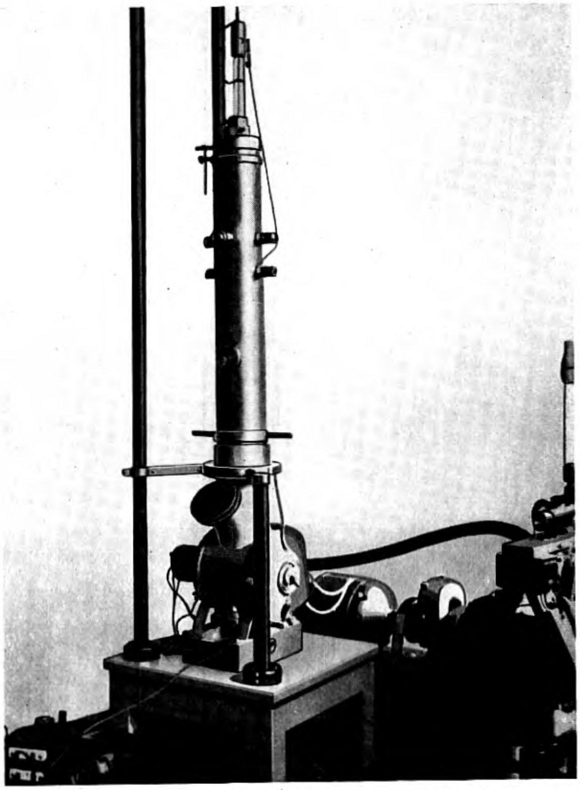


Fig. 17—Westinghouse-Norinder, 1928.[17]

cathode rays that has come through the thin window of metal or of cellophane.

We have, then, three general classes of cathode ray oscillographs: First those that resemble the original Braun tube, still in limited use, comprising a glass tube with a fluorescent screen, and a relatively high operating voltage; second the tubes of the Dufour type that I have just described, with direct recording on the photographic plate or

[16] Knoll, Max, *Zeits. f. tech. Phys.*, **10**, p. 28, 1929.
[17] Norinder, H., *A. I. E. E. Trans.*, **47**, p. 446, 1928.

film; and third, tubes with a hot cathode, with a relatively low operating voltage.

Of this last type is the Western Electric No. 224 Cathode Ray Oscillograph. Since this is the tube with which I have had more direct

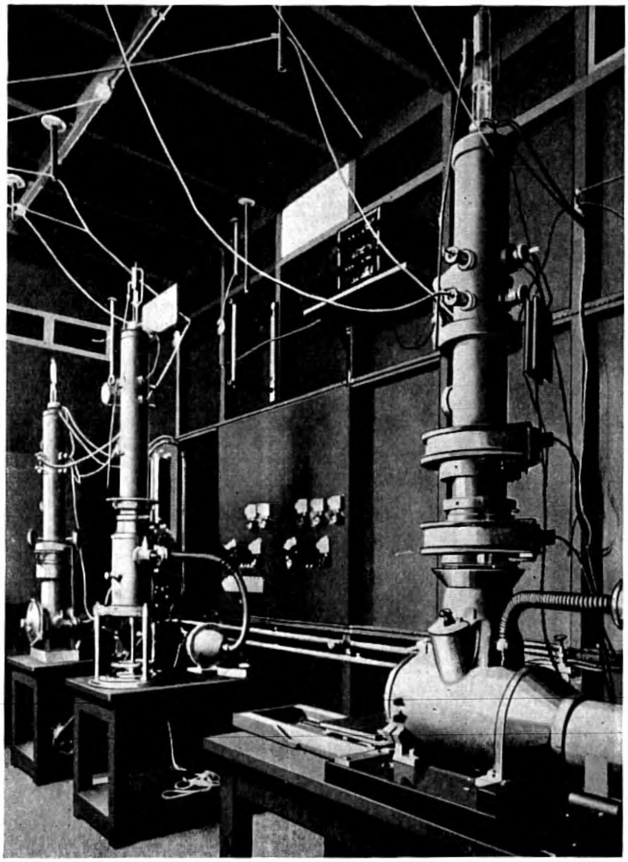


Fig. 18—Norinder, 1930—group of oscillographs.[18]

contact I wish to discuss some problems of its development and operation.

The design of the tube shown in Fig. 19 is now fairly familiar,* and

[18] Norinder, H., *Zeits. f. Phys.*, **63**, p. 672, 1930.

* Since the paper was read the tube has been altered in certain respects. The glass enclosure for the filament has been replaced by one of metal, and the anode and deflector plates are supported on machined insulating blocks. These changes make the structure more rugged and insure more nearly perfect alignment of the parts. The end of the bulb has been changed so as to present a cylindrical surface instead of a spherical one, thus permitting more intimate contact between the fluorescent screen and a photographic film when contact photographs are made.

Fig. 19—Western Electric tube.

I shall describe it only as I describe the reason for the various features.

First we wanted a convenient tube to operate on the batteries of an ordinary vacuum tube. The thermionic filament cathode was therefore required. The anode is a metal tube placed a short distance from the cathode and between them is a metal disc with a perforation through which the electrons pass to the anode. These electrodes are represented respectively by the letters *C*, *A* and *D* in Fig. 20.

The electrons flow from the cathode to the inside of the anode and some of them pass through and form the electron beam. Now, for

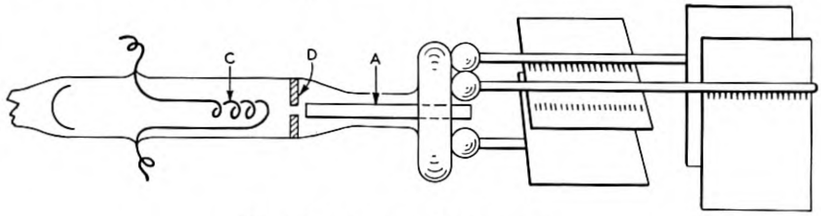


Fig. 20—Diagram of electron gun.

reasons to be mentioned presently, there is some gas in the tube and this puts two requirements on the structure of the electron gun. First, the gas would permit considerable ionization in the tube if the electrons were permitted to pass into it. Most of the current would go around the disc and to the outside of the anode. The cathode and anode are therefore enclosed in a small tube, having dimensions less than the mean free path of electrons in the gas so that no appreciable ionization can build up. There is some ionization, however, in the space between the cathode and anode, and the positive ions produced tend to bombard the filament. If the filament is directly exposed to this bombardment the oxide coating is worn off, as in a sand blast, in a matter of two or three hours. The filament is therefore wound in the shape of a helix which is mounted coaxially with the anode and the perforation in the disc so as to be out of the direct path of the ions. In this way the filament is made to last several hundred hours of operation.

This completes the internal parts of the electron gun. Externally it has mounted on it two pairs of deflector plates that control the direction of the electrons after they leave the gun. In order to prevent any large difference of potential between the anode and deflector plates, one plate of each pair is connected directly to the anode, and only the other plate has the variable potential impressed on it. As for size and separation, the plates are designed to give maximum sensitivity for a given full deflection. The sensitivity is about 1 mm.

deflection for each volt applied to the deflector plates, or 1 mm. for each ampere turn in a pair of small coils placed outside of the tube. These figures are for the normal driving potential of 300 volts.

The flattened end of the bulb, on which the electron beam impinges, is covered with a fluorescent material. The powder is a mixture of zinc orthosilicate and calcium tungstate, both specially prepared for fluorescence. The zinc silicate produces a green light of high visibility and the tungstate a blue light of high photographic activity, so that with the mixture the same tubes can be used efficiently for both visual and photographic observations.

As said before, there is some gas in the tube. One purpose of the gas is to produce a small amount of ionization in the tube which prevents any unduly large charges from accumulating on the glass walls and screen. Electrons deposited on the glass are neutralized by positive ions produced in the gas. An electron current equal to the current in the beam drifts back through the gas to the anode. The chief result of this drift of electrons is that a negative space charge is formed in the tube which decreases the speed of the electrons before they strike the screen as if the driving potential had been lowered by about 50 volts.

The other and more important purpose of the gas is to bring the electrons of the beam to a sharper focus at the screen. The beam is diffuse for two reasons, first because it is originally divergent, and secondly because of the natural electrostatic repulsion that tends to force the electrons apart. The gas serves to overcome these actions in the following way: As the beam of electrons travels down the length of the tube, some electrons collide with atoms of the gas and separate the atom into an electron and a positive ion. The impact of the electron does little to displace the massive positive ion from the position it temporarily occupies while two electrons are immediately shot out of the path. The result is a column of positive ionization down the length of the beam, with a negative space charge surrounding it. This produces a radial electrostatic field which tends to bend the path of the outer electrons of the beam inward toward the centre. The magnitude of this action depends on the degree of the differential ionization. This, again, is the greater the higher the gas pressure and the greater the current in the beam. The gas pressure must be low enough so that the larger fraction of the electrons reach the screen, and then the current in the beam must be such as to produce the desired focusing action. The heavier the ions the lower can the pressure be. The condition for a focus then involves the kind and pressure of gas, the speed of the electrons, the current in the beam and the length of the tube. In the 224 tube, with argon at the pressure of .01 mm. the focusing occurs at about 20 microamperes in the beam. Less current makes a large unfocused spot; more current produces a focus before the screen is reached, with consequent spreading of the spot again. Hence it is that the spot on the fluorescent screen is focused by adjusting the heating current of the cathode.

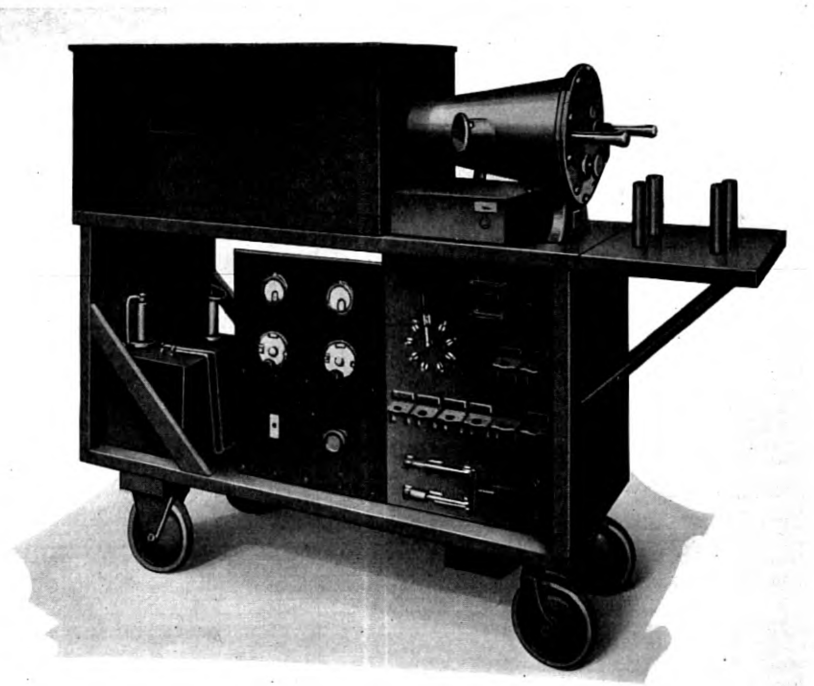


Fig. 21—General Electric, 1928.[19]

Besides focusing the electrons and preventing the accumulation of charges in the tube, the gas plays various other roles. One very curious effect of the gas is that it decreases the sensitivity of the tube at small deflecting voltages. When uniformly varying voltage is applied to a pair of deflector plates so as to move the spot across the screen, the spot seems to hesitate for an instant at the centre of the screen, appearing to be brighter there. The effect has always been observed but particular attention has recently been given to it by Professor Bedell.[20] The explanation has to do with space charge be-

[19] Lee, E. S., *G. E. Rev.*, **31**, p. 404, 1928.

[20] Bedell, F. & Kuhn, J., *Phys. Rev.* **36**, p. 993, 1930.

tween the deflector plates. The beam of electrons produces slowly moving positive ions and electrons in the gas along its path between the plates. When a voltage is applied to the plates, the positive ions travel from the beam toward the negative plate and an equal number of the electrons travel toward the positive plate. The space charge set up by the electrons and ions produces an electric field opposing that created by the applied voltage. The greatest space charge occurs near the negative plate where the sluggish positive ions flow. The space midway between the plates remains nearly field free and there is little deflection of the beam until the voltage is greater than that at which all of the ions produced are drawn to the plate. Calculation agrees with observation that this voltage is 2 to 3 volts on either side of zero.

Having now described the operation and structure of some of the oscillographs, I should like to say something more about their uses.

The cathode ray oscillograph is essentially a curve tracer that plots out in rectangular coordinates the relation between two quantities represented by the fields between the deflector plates. Often one of these quantities is time, as in the ordinary moving mirror oscillograph, while the other is some electrical quantity. We then say that we plot a wave shape. We must then have a way of making the spot move at a uniform rate. One of the simplest and most reliable ways of producing a linear time axis,[21] at least for low voltage tubes, employs two thermionic tubes to control the charge in a condenser as shown in Fig. 22. One tube, a simple two-electrode tube T_1, limits the charging

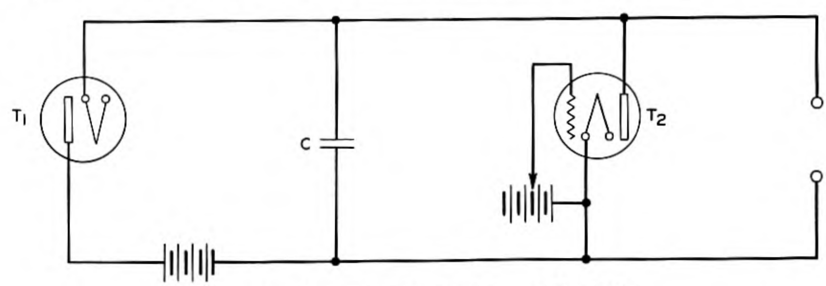


Fig. 22—Thermionic tube "sweep circuit."

current of the condenser C so that the voltage across the condenser rises linearly with time according to the equation $V = Cit$. The second tube T_2 is filled with gas and has the property of passing current only when the voltage across it reaches a certain value, which in turn is controlled by the grid potential. When the condenser volt-

[21] A. L. Samuel, *Rev. Sci. Inst.*, **2**, p. 532, 1931.

age reaches this value the tube operates to discharge the condenser suddenly, and then the uniform charging process begins over again. The condenser voltage applied to one pair of plates of the oscillograph

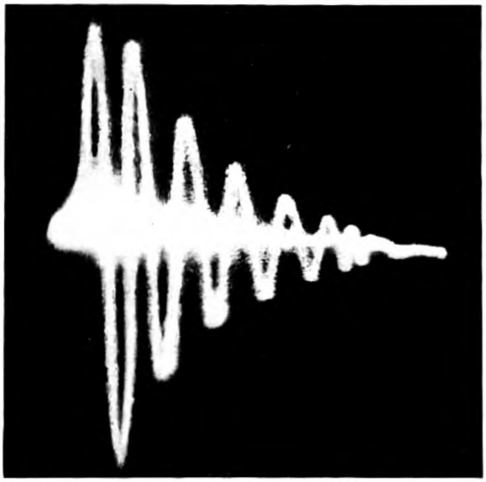


Fig. 23—Discharge of condenser through inductance.

tube makes the spot travel over the screen at a uniform rate in one direction, returning much faster in the other direction. This process may be repeated once a second, or many thousand times a second, depending on the frequency of the wave that is being studied.

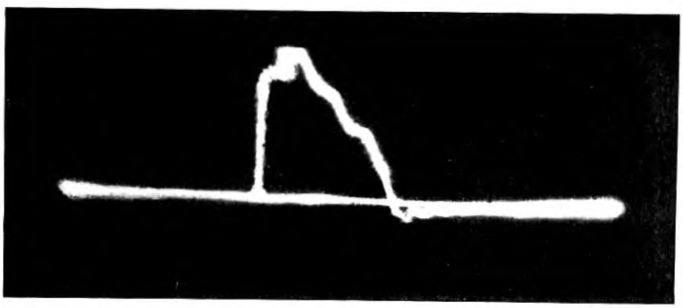


Fig. 24—Discharge of condenser through chattering contact.

In Figs. 23–29 are shown a number of records plotted on a time basis. Figs. 23–25 were made with the Western Electric tube, Figs. 26–27 with the Dufour oscillograph, and Figs. 28–29 on an oscillograph of the type of Rogowski.

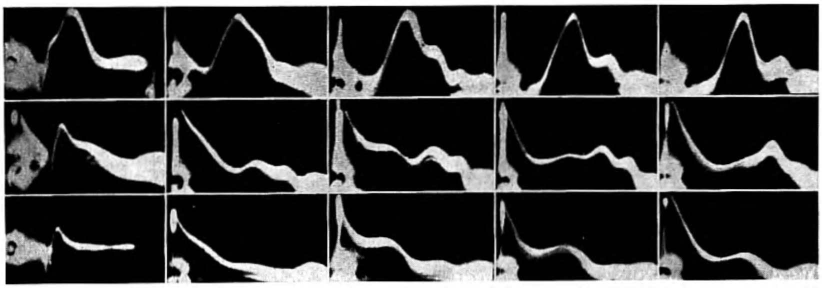


Fig. 25—Action current of a stimulated frog-nerve.[22]

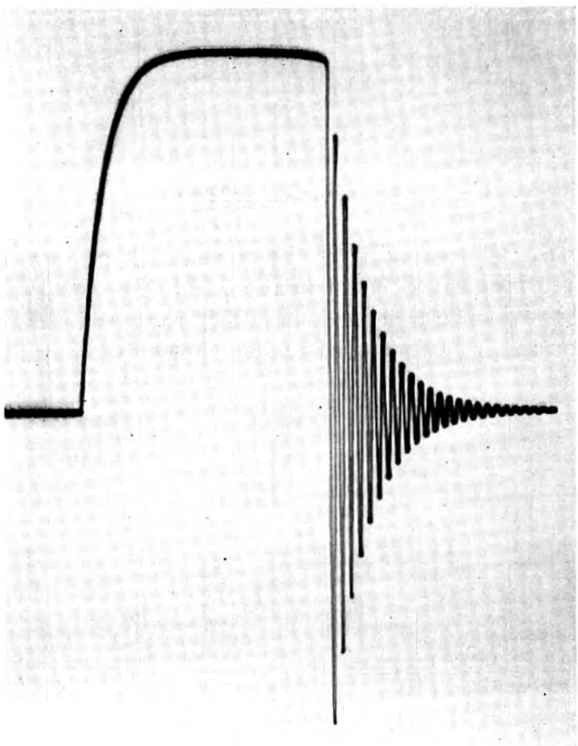


Fig. 26—Make of current through direct contact and break through tuned circuit—Dufour.[14]

[22] Gasser and Erlanger, *Am. Jl. Physiol.*, **73**, p. 613, 1925.

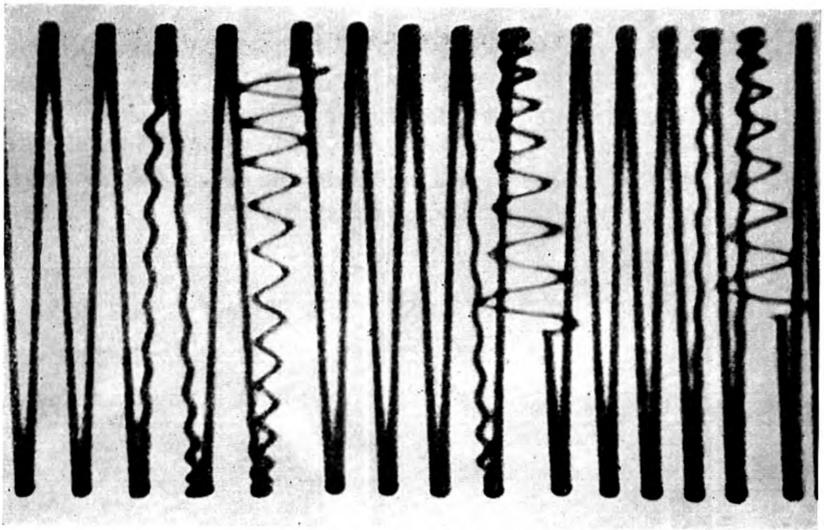


Fig. 27—Wave shape at 8,500,000 cycles—Dufour.[14]

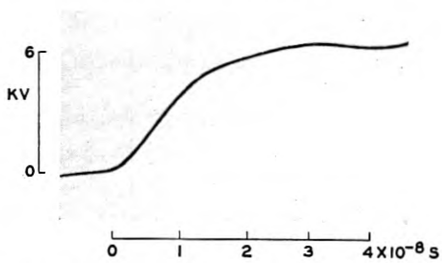


Fig. 28—Front of a voltage wave traveling on a conductor.[23]

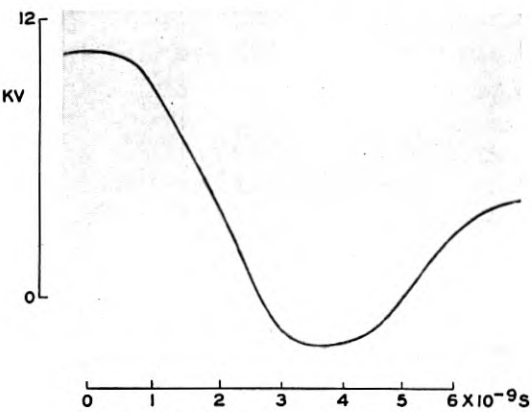


Fig. 29—Initiation of spark in a gas.[24]

[23] Krug, W., *E. T. Z.*, **51**, p. 605, 1930.
[24] Krug, W., *Zeits. f. techn. Phys.*, **11**, p. 153, 1930.

Another way to use the tube is to have it plot the relation between two quantities irrespective of time. As a simple case we may take the current *vs.* voltage curve of a resistance through which an alternating current is flowing. The spot travels back and forth along a straight line, the slope of which measures the reciprocal of the resistance (Fig. 30a). If inductance is added to the resistance the spot

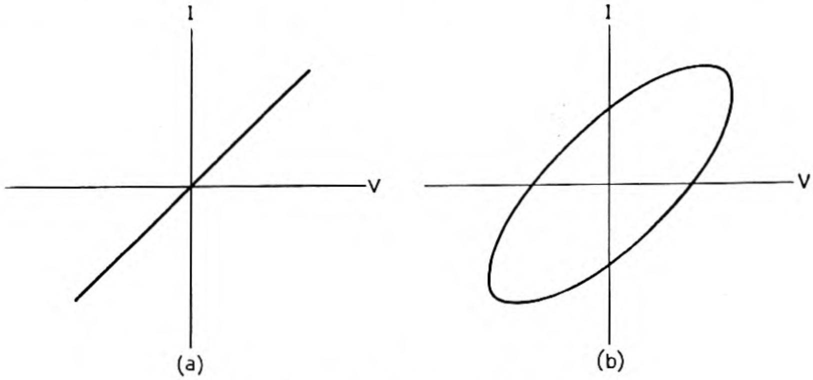


Fig. 30—Diagrams of current voltage curves.

does not come down on the same line it went up; an ellipse results, the spread of which tells us the amount of the inductive impedance (Fig. 30b). This method of operation has a wide variety of applications. Instead of the resistance, for instance, we may have a gas discharge device of which we want to know the properties. One of the applications of this method is the production of hysteresis curves of the ferromagnetic materials [25] (Fig. 31). Fig. 32 illustrates the application of the method to the study of distortion in an amplifier.

Suppose we apply to each pair of deflector plates a voltage from each of two different oscillators. If the oscillators make exactly the same number of vibrations per second, then the pattern on the tube remains stationary, but if the frequency of the oscillators differ ever so little the pattern goes through gradual changes according to the different phase relations. This is one of the most sensitive means we have for comparing and calibrating accurate oscillators, and we may call it the Lissajous figure method. Fig. 33 shows the appearance of some of these stationary Lissajous patterns.

Another method of comparing the frequency of oscillators has been called the gear-wheel method. In this method the voltage from the low frequency source is split into two equal components 90° apart in

[25] Johnson, J. B., *Bell System Technical Jl.*, **8**, p. 286, 1929.

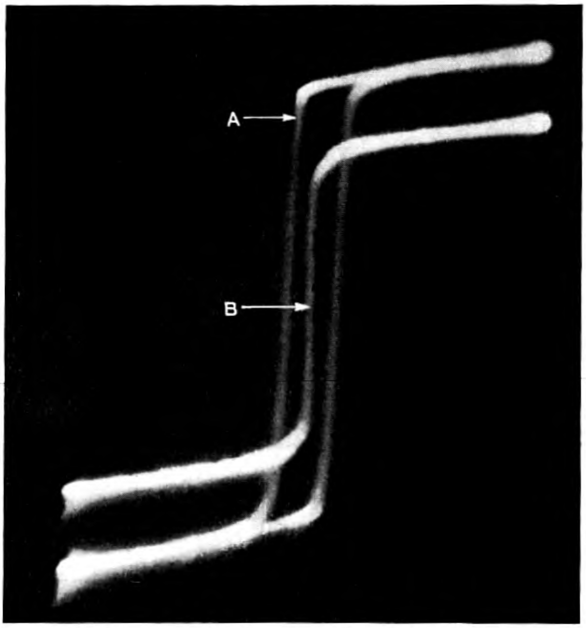


Fig. 31—Hysteresis curves (*a*) of iron, (*b*) of permalloy.

phase. These two voltages are applied to the deflector plates, producing a stationary circle on the screen. The voltage from the oscillator of higher frequency is introduced in the circuit between the cathode and anode of the tube so that the sensitivity of the tube is varied in accordance with the higher frequency. The circle is distorted into a gear-wheel shape as shown in Fig. 34, provided that the

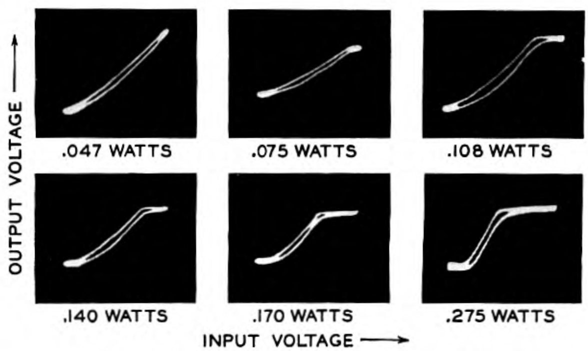
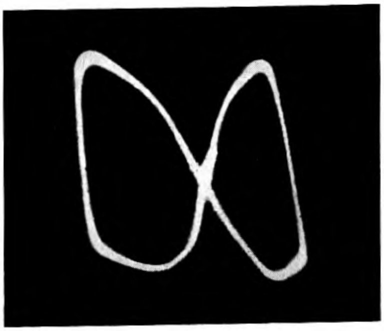
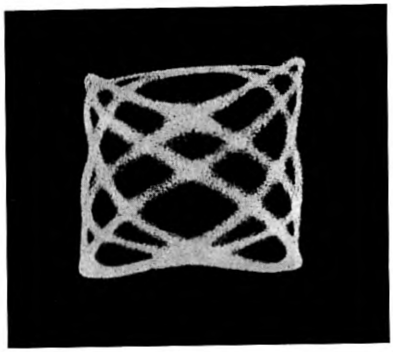


Fig. 32—Distortion in vacuum tube amplifier.[26]

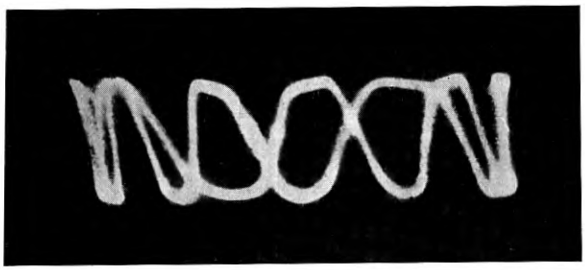
[26] Willis and Melhuish, *Bell System Technical Jl.*, **5**, p. 573, 1926.



(*a*) Ratio 2 : 1.



(*b*) Ratio 5 : 4.



(*c*) Ratio 8 : 1.

Fig. 33—Frequency comparison.

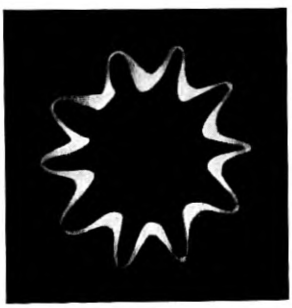


Fig. 34—"Gear wheel" frequency comparison, ratio 10 : 1.

higher frequency is an exact multiple of the lower. If the frequency ratio is not a rational number the gear-wheel rotates, showing the amount of the lack of synchronization.

Two interesting applications to the study of radio telephony are illustrated in Figs. 35 and 36. Fig. 35 shows two characteristics of a transatlantic radio channel, the tube plotting the amplitude of the received signal at a number of modulating frequencies. In Fig. 36 are plotted the magnitude and direction of static crashes as they were shown on the screen of the oscillograph tube.

For demonstration purposes the tube may be used to determine the value of the ratio of charge to mass of the electron, designated by e/m. The classical method is obvious from the derivation of the sensitivity of the tube given above. Especially is this so if we can assume the two fundamental relations $\frac{1}{2}mv^2 = eV$; $mv = eHR$. The method is

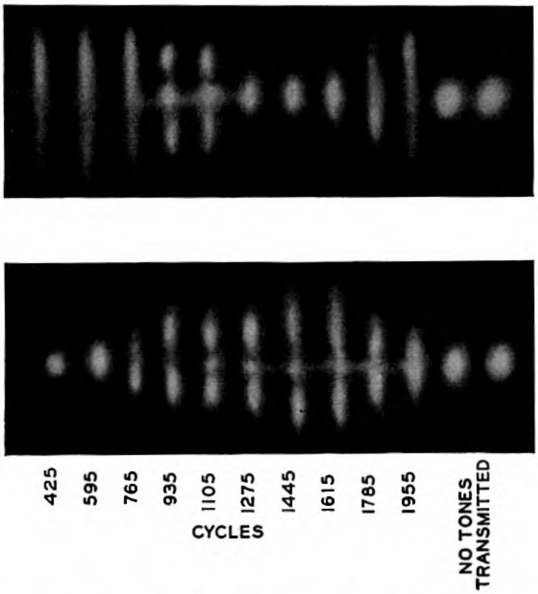


Fig. 35—Transmission characteristic of transatlantic short-wave radio channel.[27]

subject to some error because the extent of the deflecting fields cannot be exactly determined, and because of the space charges in the tube.

A more accurate method of arriving at the value of e/m is due to H. Busch.[28] A long solenoid carrying constant current creates a uniform magnetic field in the tube parallel to its axis. The electrons travel in spirals in this field in such a way that when the field has one of a series of values the electrons are focused on the screen. These critical values of the magnetic field are given by the equation

$$H = \frac{2\pi}{L}\sqrt{\frac{2m}{e}}\,V\,;$$

[27] Potter, R. K., *Inst. Radio Eng.*, **18**, p. 581, 1930.
[28] Busch, H., *Phys. Zeits.*, **23**, p. 438, 1922.

where $n = 1, 2, 3.$—etc., L is the length of the beam and V the driving potential.

In this method there is less disturbance from gas focusing and space charge if the filament current is made as low as will give a still visible spot with the magnetic focusing.

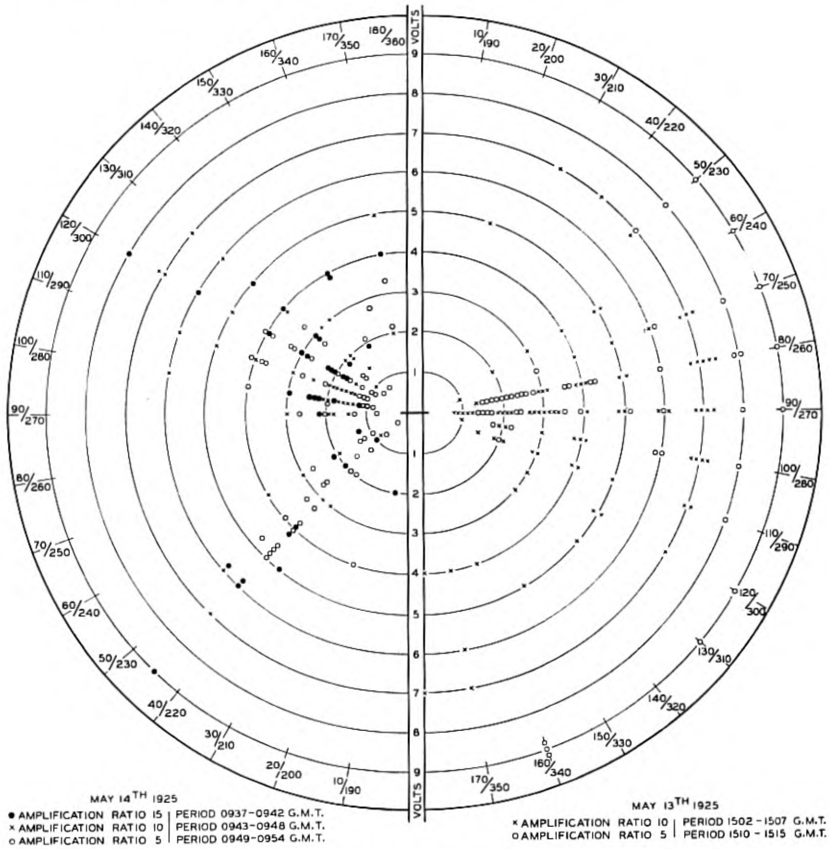


Fig. 36—Azimuthal distribution of atmospherics.[29]

From the beginning the cathode ray oscillograph was recognized as a tool of great promise. Its use was limited, however, by the difficulties of maintaining a constant degree of vacuum and of providing a suitable source of high voltage, as well as by the general bulk and clumsiness of the apparatus. We have seen that in the nearly thirty-five years since the first invention of the tube, many improvements have been made both in the structure of the tube and in the methods

[29] Watson Watt and Herd, *Jl. I. E. E.*, **64**, p. 611, 1926.

of operation so that the handicaps under which the early apparatus was used have been very largely eliminated. The introduction of the low voltage, high sensitivity tube of moderate cost expanded the use of cathode ray oscillographs rapidly, until now they are used in almost every laboratory where high frequency measurements are made. In the Bell System alone more than a hundred of these instruments are in constant use for research and for the control and calibration of manufactured products. Originally an intractable and little used device, the cathode ray oscillograph has become an almost universal scientific and industrial tool.

The Operation of Vacuum Tubes as Class B and Class C Amplifiers *

By C. E. FAY

A simple theoretical development of the action of a vacuum tube and its associated circuit when used as a Class B or Class C amplifier is given. An expression for the power output is obtained and the conditions for maximum outputs are indicated. The way in which the tuned plate circuit filters out the harmonics in the pulsating plate current wave is illustrated by an hypothetical example. A set of dynamic output current characteristics is developed graphically from a set of static characteristics. The Class B dynamic curves are found to give a better approximation to a straight line than the Class C curves because of a reversed curvature which appears at the lower ends. It is pointed out that the screen grid tube should function similarly to a high μ three-element tube in this type of operation. Experimental dynamic characteristics of a three-element tube, Western Electric 251-A, and of a screen grid tube, Western Electric 278-A, of identical dimensions are shown which verify the theoretical results. The screen grid tube gives about the same output and efficiency as the three-element tube, but its dynamic characteristic tends to bend more rapidly at the upper end.

Introduction

THE majority of modern radio telephone installations in this country are designed for modulation at a low power level and amplification of the modulated carrier, so as to obtain up to 100 per cent modulation in the output stage. As far as the writer is aware there seems to be no very comprehensive material dealing with this phase of vacuum tube operation available from past publications. A treatment of the operation of such amplifiers seems particularly desirable from the standpoint of the design of vacuum tubes for such service.

Some of the fundamental considerations regarding Class B or C operation were given by Morecroft and Friis,[1] and later, a more complete analysis of power oscillators by Prince.[2] Both of these, however, were primarily concerned with the attainment of steady output at high efficiency. Other papers [3, 4, 5] of more recent date have touched somewhat upon the subject.

* To appear in *Proc. I. R. E.*, March, 1932.

[1] J. H. Morecroft and H. T. Friis, "The Vacuum Tube as a Generator of Alternating Current Power," *Transactions A. I. E. E.*, Vol. 38, No. 2, Oct., 1919.

[2] D. C. Prince, "Vacuum Tubes as Power Oscillators," *Proc. I. R. E.*, Vol. 11, Nos. 3, 4, 5, June, Aug., Oct., 1923.

[3] A. A. Oswald and J. C. Schelleng, "Power Amplifiers in Transatlantic Radio Telephony," *Proc. I. R. E.*, Vol. 13, No. 3, June, 1925.

[4] E. E. Spitzer, "Grid Losses in Power Amplifiers," *Proc. I. R. E.*, Vol. 17, No. 6, June, 1929.

[5] Y. Kusunose, "Calculation of Characteristics and the Design of Triodes," *Proc. I. R. E.*, Vol. 17, No. 10, October, 1929.

This paper will deal particularly with the type of amplifier used for the amplification of the modulated carrier. It will be assumed that a linear relation between input voltage and output current is the desired characteristic of such amplifiers and that the more nearly this relation is attained, the less will be the distortion produced.

An approximate graphical method for the calculation of the dynamic output characteristic from the static characteristics of a tube is outlined which is capable of considerable accuracy. The exciting voltage is taken to be a sinusoidal voltage of varying amplitude, and only the fundamental component of the output current is considered. The very important question of the distortion introduced by the non-linearity of the characteristic and by the resonance effect of the output circuit, as well as the question of the suppression of harmonics in the antenna circuit, is considered to be beyond the scope of this paper.

Theoretical Development

Class B [6] amplifiers have been defined as those which operate with a negative grid bias such that plate current is practically zero with no excitation grid voltage, and in which the power output is proportional to the square of the excitation voltage.

Class C amplifiers have been defined as those which operate with a negative grid bias more than sufficient to reduce the plate current to zero with no excitation grid voltage, and in which the output varies as the square of the plate voltage between limits.

There is actually very little distinction between the two types as the fundamental principles of operation are the same in that the plate current flows in pulses and becomes zero during part of the cycle, the Class C type being merely the case where the duration of the pulses is shorter. For the purposes of this paper, a Class B amplifier shall be regarded as one in which the grid bias is either just sufficient or is not sufficient to reduce the plate current to zero with no excitation grid voltage, and a Class C amplifier as one in which the grid bias is more than sufficient to reduce the plate current to zero with no excitation grid voltage.

Let us consider the plate circuit of a Class B or C amplifier to be represented schematically in Fig. 1. We will omit the grid circuit and assume that the excitation of the grid merely varies the internal tube resistance, R_p, which it does in effect. If at the start, the grid is so biased that no plate current flows, the condenser C_0 will charge up to a potential E_b as indicated in Fig. 1. Let it be assumed that C_0 is of sufficiently large capacity that it presents negligible impedance at the

[6] 1931 Standardization Report, Year Book of the I. R. E., 1931, p. 71.

frequency of operation, or that the time constant of C_0 and R_0 is large compared to the time of one cycle of the operating frequency. In this event, then, the voltage across C_0 will remain constant at E_b during a complete cycle. Let it also be assumed that the choke coil L_b is of sufficient inductance that the current I_b is maintained constant throughout the cycle.

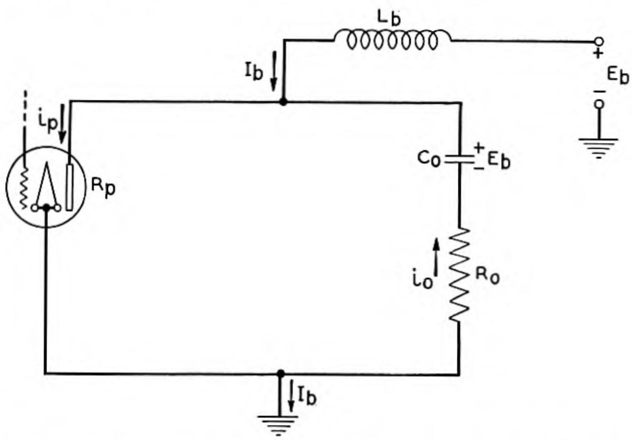


Fig. 1—Schematic of amplifier plate circuit with resistance load.

Then by applying Kirchoff's laws to the circuit, Fig. 1, remembering the assumptions regarding C_0 and L_b, we find the following relations must hold at any instant:

$$E_b = i_pR_p + i_0R_0 \tag{1}$$

or

$$e_p = E_b - i_0R_0 \tag{2}$$

and

$$i_0 = i_p - I_b, \tag{3}$$

also

$$I_b = \text{Average of } i_p \text{ over 1 cycle}, \tag{4}$$

since the average current through C_0 must be zero.*

Then let it be assumed that the grid of the tube is excited and biased in such a manner that the plate current, i_p, will vary sinusoidally as illustrated in Fig. 2. When the resistance R_p is at its minimum value, i_p will be a maximum and will be the sum of I_b and i_0, [(1) and (3)]. Also e_p will be a minimum, (2); see point 1, Fig. 2. As R_p is increased by the grid potential going in the negative direction, i_p will be reduced

* See Table of Symbols at end of this paper.

as also will i_0. At the instant when

$$i_p = I_b, \qquad i_0 = 0,$$

as illustrated by point 2, Fig. 2. Then as R_p is further increased, i_p

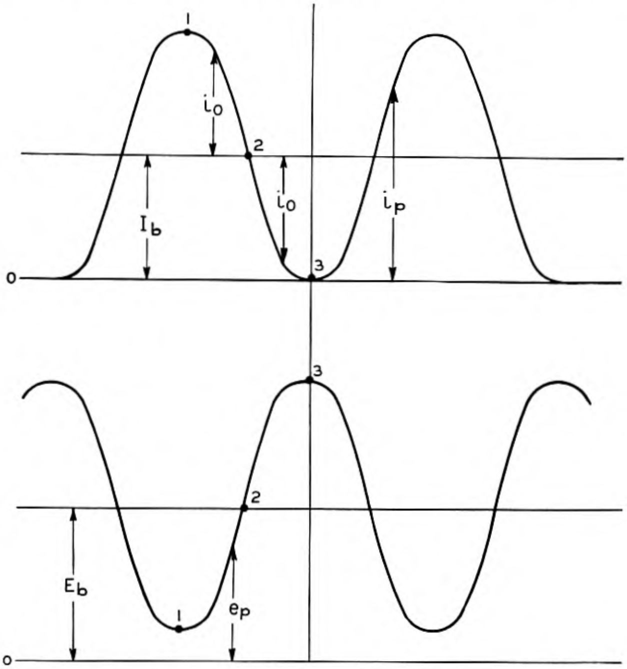


Fig. 2—Plate current and plate voltage relations for sinusoidal plate current in circuit of Fig. 1.

becomes less and i_0 starts to increase in the negative direction until i_p is zero at cut-off;

$$R_p = \infty, \qquad i_0 = -I_b \text{ from (3).}$$

Then from (2) and the above,

$$e_p = E_b + I_b R_0, \tag{5}$$

which is the peak value of e_p. Thus the power in R_0 is $i_0^2 R_0$ averaged over a cycle and the power dissipated in the tube is $i_p^2 R_p$ averaged over a cycle.

In actual operation, the plate current is not sinusoidal but goes from zero to peak value and back to zero in about half the cycle and remains cut off during the rest of the cycle, as illustrated by Fig. 4A. Also, instead of being a pure resistance, the output circuit consists of a tuned tank circuit, Fig. 3, into which resistance is introduced either directly or by coupling in some way. Thus the impedance of the output circuit will be a resistance to the fundamental frequency only. However, the same general principles will apply in this case as in the case

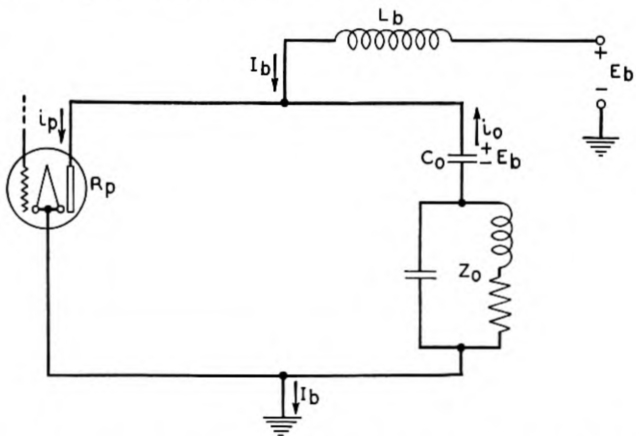


Fig. 3—Schematic of amplifier plate circuit with tuned output impedance.

of the sinusoidal plate current and the pure resistance circuit. The equations will be of the same form except that Z_0 must be substituted for R_0 as follows:

$$E_b = i_p R_p + i_0 Z_0, \tag{1A}$$

$$e_p = E_b - i_0 Z_0, \tag{2A}$$

$$i_0 = i_p - I_b, \tag{3}$$

$$I_b = \text{Average of } i_p \text{ over 1 cycle.} \tag{4}$$

In this case it must be remembered that the wave of i_p, and hence i_0, instead of being a simple sine wave, is a more complex wave consisting of a fundamental frequency and numerous harmonics. Also i_0Z_0 is the sum of all such components multiplied by the respective impedances presented to them taken in their proper phases.

Considering the wave form of i_0 shown in Fig. 4B, which is readily obtainable in practice, it should be evident from inspection that it can be considered a cosine wave containing odd and even cosine terms and may be expressed by

$$i_0 = I_1 \cos \omega t + I_2 \cos 2\omega t + I_3 \cos 3\omega t + I_4 \cos 4\omega t + \cdots + I_n \cos n\omega t.$$

By means of an harmonic analysis of this wave for the numerical values indicated in Fig. 4B, the coefficients were found to be approximately as follows:

$$I_1 = 0.96, \quad I_2 = 0.543, \quad I_3 = 0.140, \quad I_4 = -0.07, \quad I_5 = -0.105, \quad I_6 = -0.043.$$

The voltage produced across the output circuit by the wave, i_0, then will be the sum of the voltages produced by each one of the com-

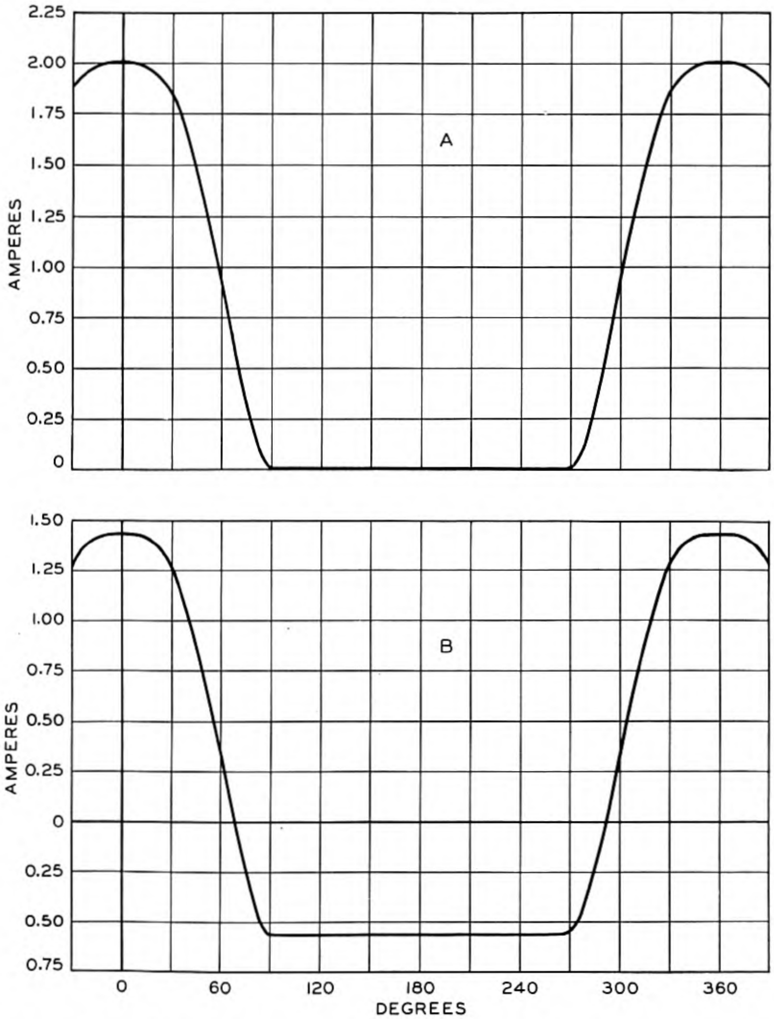


Fig. 4—*A*. Pulsating plate current wave, i_p. *B*. Alternating component of plate current, i_0.

ponents of i_0. Let it be assumed that the reactances of the output circuit are 300 ohms each at fundamental frequency and that sufficient resistance has been inserted in the inductive branch to make the impedance to fundamental frequency 2000 ohms resistance. This will require about 45 ohms in series with the inductance. At second harmonic frequency, then, the impedance of the inductive branch in complex notation is $45 + j600$ and that of the capacity branch $0 - j150$. Without introducing much error in the result we may as well write $+ j600$ and $- j150$, which in parallel give $- j200$. At third harmonic frequency we have $+ j900$ and $- j100$ which in parallel give $- j112.5$. At fourth harmonic frequency we have $+ j1200$ and $- j75$ which in parallel give $- j80$.

Thus the voltage produced by the fundamental will be

$$e_1 = 0.96 \times 2000 \cos \omega t = 1920 \cos \omega t,$$

and that produced by the second harmonic will be

$$e_2 = 0.54 \times 200 \cos (2\omega t - 90°) = 108 \cos (2\omega t - 90°),$$

and that produced by the third harmonic will be

$$e_3 = 0.14 \times 112.5 \cos (3\omega t - 90°) = 15.75 \cos (3\omega t - 90°),$$

and that produced by the fourth harmonic will be

$$e_4 = - 0.07 \times 80 \cos (4\omega t - 90°) = 5.6 \cos (4\omega t + 90°),$$

etc. Fig. 5B shows to scale the fundamental and second harmonic voltages produced, and the dotted curve gives the sum. The higher harmonic voltages are too small to show on the plot. The resultant is seen to be very little different from a sine wave. Therefore, even though the wave i_0 departs radically from a sine wave, the voltage produced across the output circuit is very nearly sinusoidal. Then the output power at fundamental frequency is

$$W_0 = \frac{I_1^2 R_0}{2}, \tag{6}$$

where I_1 is the peak value of the fundamental component of i_0, and R_0 is the effective resistance of the tank circuit at fundamental frequency.

The output power may also be expressed:

$$W_0 = \frac{(E_b - e_{pm}) K I_p}{2} \tag{7}$$

if $I_1 = KI_p$.

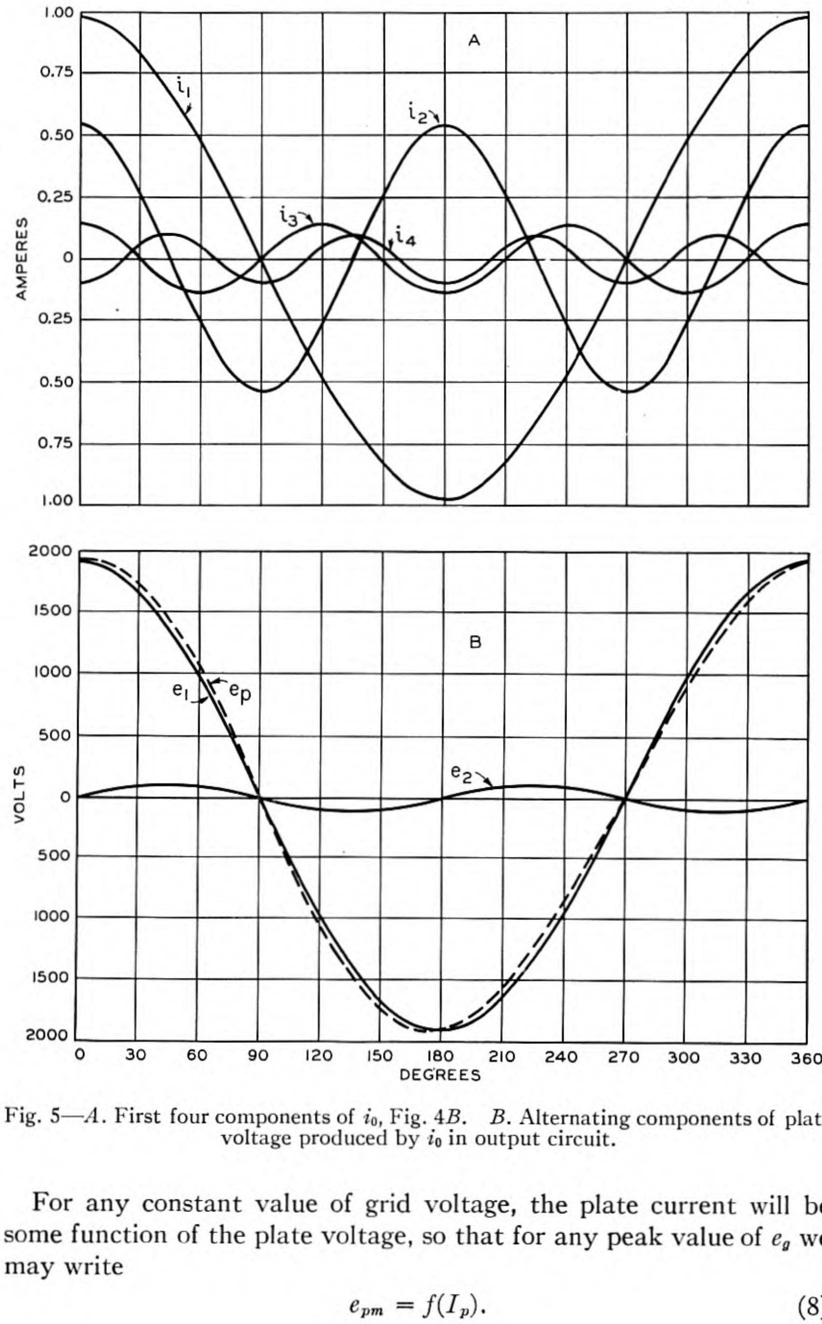


Fig. 5—*A*. First four components of i_0, Fig. 4*B*. *B*. Alternating components of plate voltage produced by i_0 in output circuit.

For any constant value of grid voltage, the plate current will be some function of the plate voltage, so that for any peak value of e_g we may write

$$e_{pm} = f(I_p). \tag{8}$$

Substituting (8) in (7) we have

$$W_0 = \frac{[E_b - f(I_p)]KI_p}{2}. \qquad (9)$$

In order for W_0 to be a maximum,

$$\frac{dW_0}{dI_p} = 0.$$

Performing this operation on (9) and solving for I_p assuming K is constant, we get

$$I_p = \frac{E_b - f(I_p)}{f'(I_p)} = \frac{E_b - e_{pm}}{r_p} \qquad (10)$$

since $f'(I_p)$ is $df(I_p)/dI_p$, which is obviously r_p, the differential plate resistance when I_p is flowing.

We may also write

$$R_0 = \frac{E_b - e_{pm}}{KI_p}. \qquad (11)$$

Substituting (10) in (11) we obtain

$$R_0 = \frac{r_p}{K} \qquad (12)$$

which gives the relation of R_0 to r_p for maximum power output.

If we have the $i_p - e_p$ curves for any tube we may approximate closely the point of maximum output at any peak grid voltage. This can be accomplished by a process of cut and try in finding where the quantity $(E_b - e_{pm})I_p$ becomes a maximum, since K remains fairly constant, depending mostly on the bias voltage and grid excitation voltage, as will be shown later. Also, for any given output impedance and peak grid voltage, we may find the output from the $i_p - e_p$ curves by cut and try by finding where

$$\frac{E_b - e_{pm}}{I_p} = KR_0 \qquad \text{[from (11)]}$$

is satisfied. It will be noticed from Figs. 4A and 5A that $I_1/I_p = 0.48$. In general, K will be near this value for well loaded conditions. The actual value depends upon the exciting voltage and grid bias voltage, and at low values of excitation it may differ considerably from the value indicated above which may render the cut and try methods outlined subject to considerable error.

In addition to finding the maximum output for a given set of conditions it is desirable to know what the shape of the curve of output current versus exciting voltage will be. The following method determines points on the output curve by the use of the static characteristic of the tube.

In Fig. 6 is shown a family of plate current curves obtained by applying a sine wave grid voltage to a three-halves power characteristic with a plate voltage consisting of a steady voltage plus a sine wave 180° out of phase with the grid voltage. The different curves show the relative shapes obtained by variation of the grid bias so that the portion of the cycle during which plate current flows is varied. It can

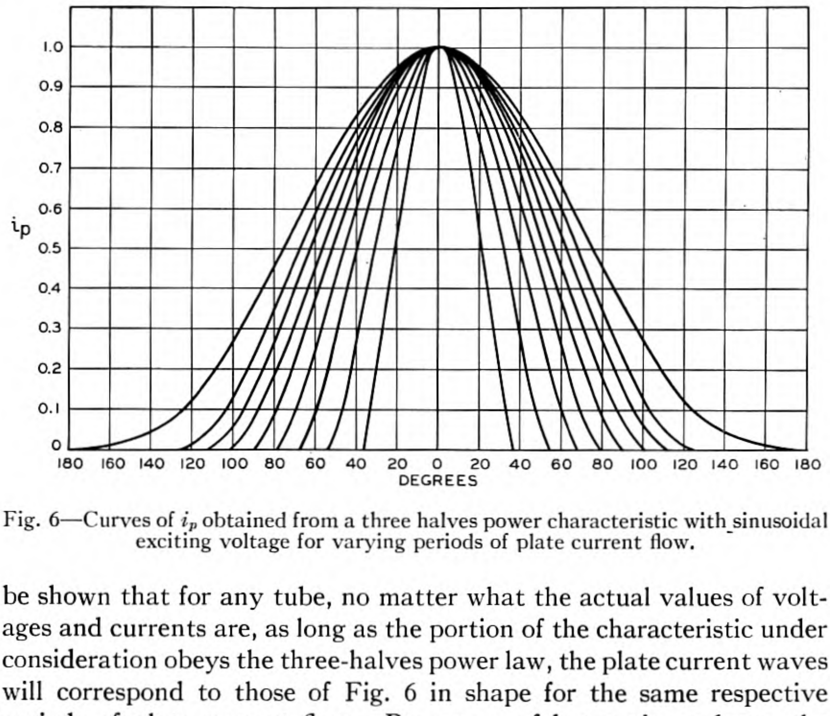


Fig. 6—Curves of i_p obtained from a three halves power characteristic with sinusoidal exciting voltage for varying periods of plate current flow.

be shown that for any tube, no matter what the actual values of voltages and currents are, as long as the portion of the characteristic under consideration obeys the three-halves power law, the plate current waves will correspond to those of Fig. 6 in shape for the same respective periods of plate current flow. By means of harmonic analyses, the value of K corresponding to each of these shapes was calculated and Fig. 7 shows the variation of K with the number of degrees during which plate current flows. Fig. 8 shows a comparison of three possible shapes for 180° flow with the corresponding values of K. A is the unsaturated curve of Fig. 6. B shows a curve for which the characteristic departs from the three-halves power law at its upper end, presumably due to the effects of grid current, etc. This curve if unsaturated would have a peak value about 17 per cent higher. C shows

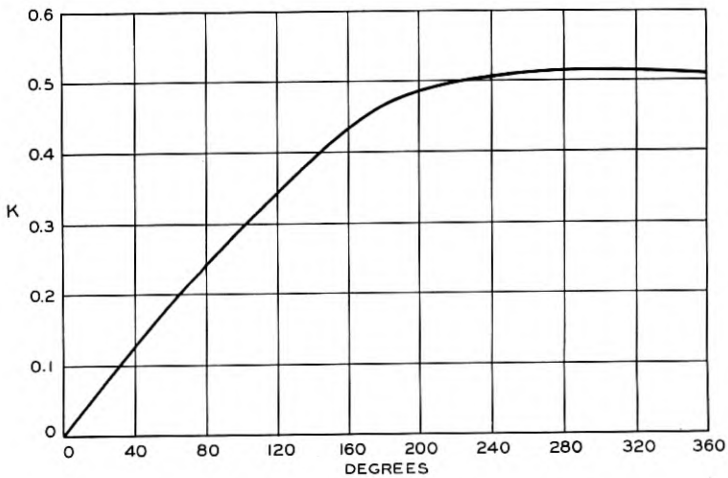


Fig. 7—Variation of K with period of flow for curves of Fig. 6.

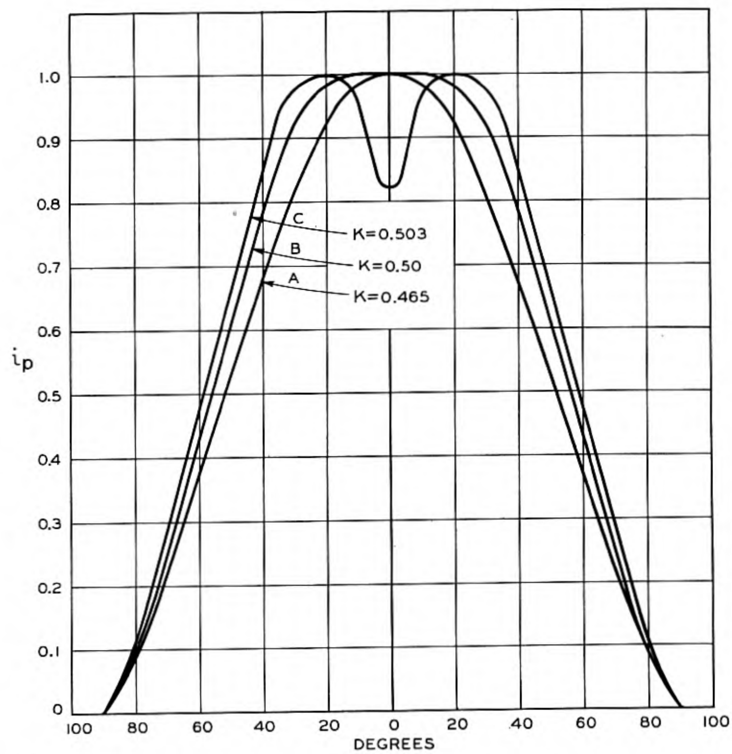


Fig. 8—Three typical shapes of plate current for 180° flow. *A*, unsaturated. *B*, slightly saturated. *C*, extremely saturated.

a curve for which the saturation is so pronounced that the grid current drawn has caused a depression.

In actual tubes the static characteristics do not follow the three-halves power law exactly, but for the purposes of this paper it is sufficiently accurate to assume that they do in the portion where grid current is not appreciable.

With the information available from Figs. 6 and 7, and the static characteristic of a tube, we should be able to plot the dynamic characteristic for any value of output impedance. For an example, let Fig. 9 represent the static characteristics of a tube (Western Electric

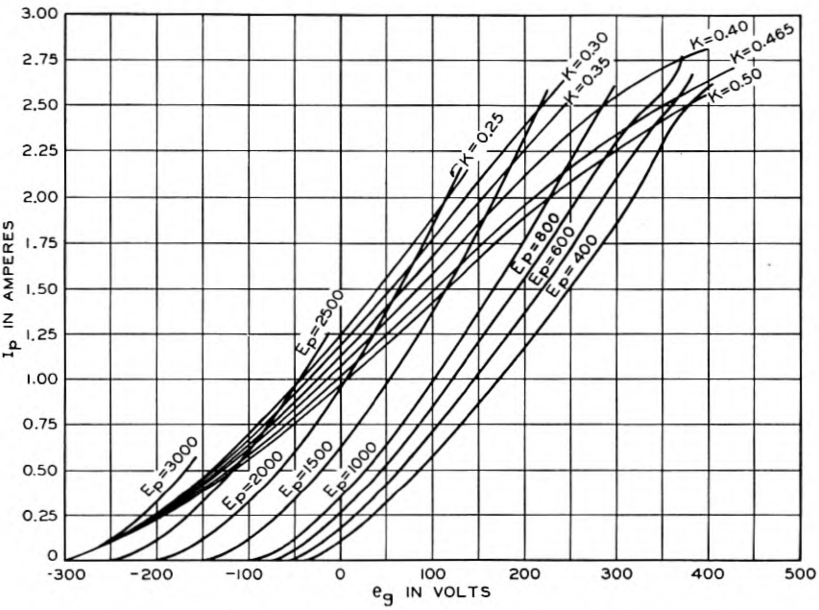


Fig. 9—Static characteristic with lines of constant K for 2000 ohms impedance. (Western Electric No. 251-A Tube.)

No. 251-A). Then for some value of output impedance, say 2000 ohms, let us plot curves of constant K on the static characteristic assuming $E_b = 3000$ volts. Points on these curves are found from

$$I_p = \frac{E_b - e_{pm}}{KR_0},$$

which is another form of (11). For example, to find where the $K = 0.5$ curve crosses the $E_p = 2000$ curve, we have $e_{pm} = E_p = 2000$, $E_b - e_{pm} = 1000$, thus

$$I_p = \frac{1000}{0.5 \times 2000} = 1.0 \qquad \text{(Fig. 9)}.$$

The cut-off point, or the potential which the grid must have to make the plate current just zero is given by

$$e_g \text{ cut-off} = -\frac{E_b}{\mu}. \tag{13}$$

If we bias the grid with this negative voltage, it may be said to be biased at cut-off, and if the grid voltage becomes more positive than this value, plate current will flow. Thus if we apply a sinusoidal exciting voltage, the plate current will always flow during 180°, or half the cycle, and for this condition $K = 0.465$ from Fig. 7, so the dynamic output current curve will be obtained by multiplying the ordinates of the $K = 0.465$ curve of Fig. 9, by 0.465. However, in the upper portion where the static characteristics depart appreciably from the three-halves power law, K will increase gradually (Fig. 8), so that it is more accurate to use slightly increasing values of K in this portion. (See table of calculations for $E_c = 300$, Fig. 10.)

$E_b = 3000$ Volts $Z_0 = 2000$ ohms

E_c Volts	Peak Excitation Volts	Peak e_g Volts	K	I_p Amp.	I_1 Amp.	$\frac{I_1}{\sqrt{2}}$ Amp.	Watts Output $\left(\frac{I_1}{\sqrt{2}}\right)^2 Z_0$	Output Current in 150 ohm Ant. Amp.
−300	100	−200	0.465	0.230	0.107	0.0756	11.5	0.276
	200	−100	0.465	0.575	0.267	0.1885	71	0.688
	300	0	0.465	1.01	0.47	0.332	222	1.218
	400	+100	0.465	1.49	0.692	0.490	480	1.790
	500	200	0.47	1.95	0.917	0.647	840	2.37
	600	300	0.48	2.30	1.07	0.756	1150	2.96
	650	350	0.49	2.45	1.20	0.848	1440	3.10
−350	100	−250	0.345	0.112	0.039	0.027	1.49	0.099
	200	−150	0.415	0.410	0.170	0.122	28.9	0.438
	300	− 50	0.435	0.810	0.352	0.249	124	0.907
	400	+ 50	0.445	1.280	0.570	0.403	324	1.470
	500	150	0.45	1.77	0.797	0.563	635	2.055
	600	250	0.452	2.20	0.995	0.702	988	2.56
	650	300	0.453	2.40	1.087	0.768	1180	2.80
−250	50	−200	0.51	0.220	0.112	0.079	12.6	0.290
	100	−150	0.505	0.375	0.190	0.134	36	0.490
	200	− 50	0.492	0.760	0.374	0.265	140	0.966
	300	+ 50	0.483	1.220	0.590	0.417	348	1.525
	400	150	0.480	1.710	0.820	0.580	674	2.120
	500	250	0.477	2.137	1.020	0.721	1040	2.635
	600	350	0.475	2.475	1.173	0.830	1380	3.030

Fig. 10—Calculation of dynamic output characteristics—No. 251-A tube.

If the grid bias is more negative than the cut-off point (i.e., if the tube is biased "below cut-off"), the plate current will not begin to flow until the grid potential has reached the cut-off point (13), and the portion of the cycle in which plate current flows will be determined by the number of degrees of the cycle in which the exciting wave is above the cut-off point. Thus in this case, K will depend on the amplitude of the exciting wave compared to the difference between the bias and the cut-off point. When the amplitude becomes large compared to this difference, the period of plate current flow approaches 180° but when the peak value of the exciting wave just reaches the cut-off point, the period of flow is zero. In this case K can vary from almost 0.465 to zero depending on the amplitude of the exciting wave, or if the exciting voltage is sufficient to produce a saturation effect in the plate current wave, K might exceed 0.465. This is the case of the Class C amplifier.

If the grid bias is more positive than the cut-off point (i.e., if the tube is biased "above cut-off"), plate current flows during 360° of the cycle until the amplitude of the exciting voltage is sufficient to reach the cut-off point. The period of flow continues to decrease until it approaches 180° as the amplitude of the exciting voltage becomes large compared to the difference between the bias and cut-off potentials. Here K is always greater than 0.465; see Fig. 7. This is the case of the Class B amplifier.

Fig. 10 outlines the calculation of dynamic characteristics of both types from the static characteristics of Fig. 9. In these calculations, the cut-off point is assumed to be constant at -300 volts and the value of K is obtained from Fig. 7 after determining the portion of the cycle in which the exciting voltage is above the cut-off point. For $E_c = -300$ of course it is always 180°. For $E_c = -350$, and a peak exciting voltage of 100 volts for example, the peak e_g will be -250 which will give 120° for the period of flow and thus $K = 0.345$ from Fig. 7. Then on the static characteristic, Fig. 9, taking $e_g = -250$, and by interpolating between the $K = 0.30$ and $K = 0.35$ curves to $K = 0.345$ we find $I_p = 0.112$. Multiplying this by K we get $I_1 = 0.0386$. Changing to r.m.s. value, squaring and multiplying by 2000 ohms, we obtain 1.49 watts for the output power. It is then determined that this would represent a current of 0.0995 ampere in the 150-ohm dummy antenna which was used in the experimental work, and which was so coupled to the output tank circuit that the impedance into which the tube was working was of the value indicated.

In the calculations of Fig. 10, the cut-off point was assumed constant at $-E_b/\mu$. However, it actually varies as $-e_p/\mu$ and will vary sinusoidally if e_p is sinusoidal. No error is introduced by this fact in the case of bias at cut-off ($E_c = -300$). However, in the other two cases an error is introduced. In the case for bias below cut-off ($E_c = -350$) the amount of error is indicated by the two points illustrated in Fig. 11. A is the case for a peak exciting voltage of 300 volts. From the static characteristics of Fig. 9 we find that e_{pm} will be about 2300 volts which will make -230 volts, the peak of the real cut-off curve ($-e_p/\mu$) and since it is a sine wave we can plot

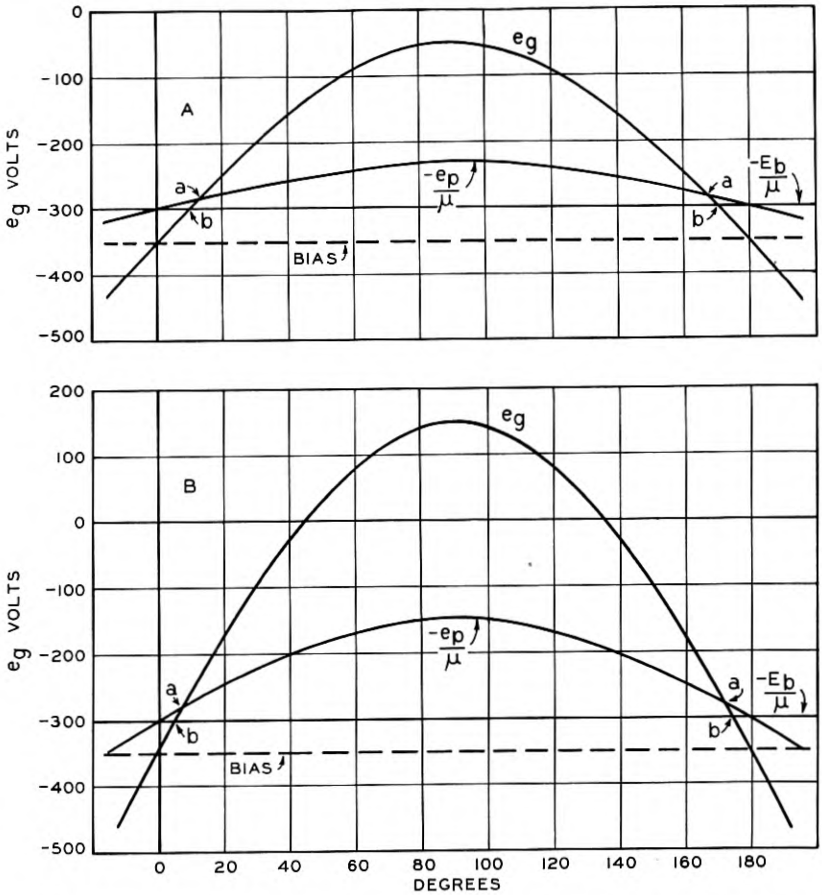


Fig. 11—Illustrating method of obtaining actual period of flow of plate current. a, acutal cut-off point, b, cut-off point assumed in calculations.

it as shown. The error in the period of flow then will be represented by the difference between where the e_g curve cuts the -300-volt line and where it cuts the $-e_p/\mu$ curve. The actual period of flow is about 156° instead of 160° as found before, which would make $K = 0.425$ instead of 0.435. B is the case for a peak exciting voltage of

500 volts and by the same procedure we find $K = 0.441$ instead of $K = 0.45$. These errors are quite small and do not alter the result appreciably in these cases.

The effect will be less pronounced, the higher the μ of the tube, and more pronounced the further the bias voltage is moved from the cut-

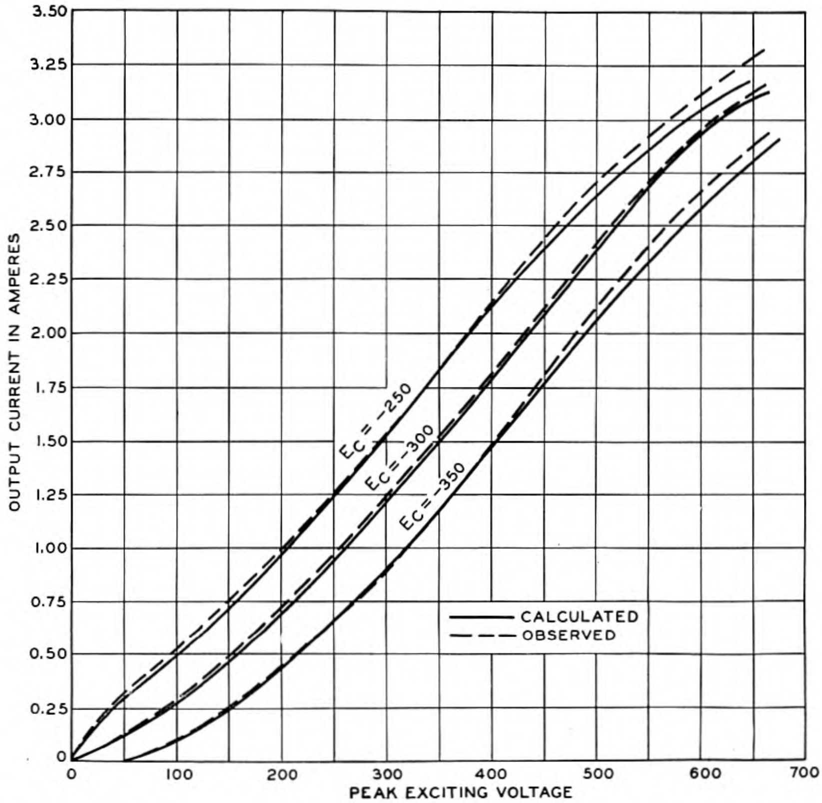


Fig. 12—Calculated and experimental dynamic characteristics of Western Electric No. 251-*A* Tube $E_b = 3000$ volts, $Z_0 = 2000$ ohms, $\mu = 10$.

off point. For extreme cases in which more accuracy is desired, K will be determined quite accurately by the second approximation method of Fig. 11. For the case of bias above cut-off where K is always greater than 0.465 the error in determining K will be considerably smaller as K does not vary greatly with the period of flow in this region; see Fig. 7. In the upper portion of the characteristic where departure from the three halves power law becomes appreciable, the increase in K due to this saturation effect which must be estimated,

Fig. 8, may be greater than the error produced by neglecting the varying cut-off. In this event it would hardly be worthwhile to use the method of Fig. 11, except in extreme cases where the bias differs greatly from the cut-off potential.

The dynamic output current characteristics calculated in Fig. 10 are shown plotted to scale in Fig. 12. It should be noted that for the case of bias above cut-off ($E_c = -250$) the curvature of the dynamic is reversed at the lower end. This will tend to allow a better approximation to a straight line for the overall curve than can be obtained for curves of bias below cut-off.

In general the highest output impedance admissible will give the straightest dynamic curve for a given tube. Comparing two tubes of equal power rating, the one with the highest mutual conductance will give the straightest dynamic characteristic with a given output impedance.

It will be noticed that the dynamic curves tend to show a saturation effect at their upper ends. This is predicted, however, from the static characteristics. If the grid excitation voltage is sufficient to allow the grid potential to approach the plate potential very closely, the electron current taken by the grid will increase rapidly. Since this current taken by the grid would otherwise have been taken by the plate, the result is a reduction in the value of i_p which causes the plate current characteristic to depart from the three-halves power law. A slight saturation effect at the upper end of the dynamic curve may help the curve to approximate a straight line more closely. The output impedance may be chosen so as to realize this advantage provided the grid becomes positive during a sufficient portion of the cycle.

The Screen Grid Tube

The foregoing theory, with a few alterations, will apply equally well to the screen grid tube. In any screen grid tube where the screening is sufficient to reduce the grid-plate capacity to the point where operation at high frequencies without neutralization is feasible, the screen voltage will determine the plate current at a given grid voltage almost entirely, the plate voltage having very little effect. In this case the cut-off point will be given approximately by

$$e_g \text{ cut-off} = -\frac{E_s}{\mu}, \qquad (14)$$

where μ here is the μ of a three-element tube with the plate in place of the screen. A more exact formula would be

$$e_g \text{ cut-off} = -\frac{E_b + \rho E_s}{\mu(1 + \rho)}, \qquad (15)$$

where μ is as defined above, and ρ is the amplification factor of a three-element tube considering the screen as the grid. However, if ρ is large compared to μ and μ is greater than 1, the cut-off point is given quite closely by (14).

In operation the screen is by-passed to ground by a large capacity so that the screen potential will remain practically constant at E_s to insure the effectiveness of the screening action. As the minimum plate potential approaches the screen potential, the screen will begin to draw appreciable current and thus cause a saturation effect in the characteristic similar to that of the three-element tube when the minimum plate potential approaches the maximum grid potential. Also, as the maximum grid potential approaches the screen potential, the grid will begin to draw current which will subtract from the plate and screen currents, thus tending also to cause a saturation effect in the characteristic. It would seem desirable, therefore, that the screen voltage be chosen somewhere between the maximum grid potential and the minimum plate potential, the actual point depending on whether grid current or screen current is the least desirable. In operation the screen grid tube should give the same type of characteristics as indicated in Fig. 12 and have some advantage from the fact that as a high μ tube it requires less driving voltage. However, the dynamic characteristic may tend to saturate sooner than it would for the equivalent three-element tube because of the tendency of both the screen and the grid to absorb current under the conditions mentioned above.

Experimental Results

In Fig. 12 along with the curves calculated from the theoretical development are shown the actual experimental curves obtained at 3000 kc for the same conditions.

The following dynamic output current and efficiency curves of the three-element tube, Western Electric 251-A, and the screen grid tube, Western Electric 278-A, measured at 3000 kc are submitted to show the effects of various factors on the dynamic characteristic. They allow a direct comparison of the three-element tube and the screen grid tube.

The curves of Fig. 13 show the dynamic output currents and efficiencies obtained from the three element tube for the conditions indicated. The upper portions of these curves show some saturation effect and they are practically identical except that they are displaced along the abscissa by approximately the differences in grid bias. Fig. 14 shows the effect of varying the output impedance for the case of bias at cut-off. The relative effect for other biases is practically the same as for the case shown.

Fig. 15 shows the dynamic output currents and efficiencies for a

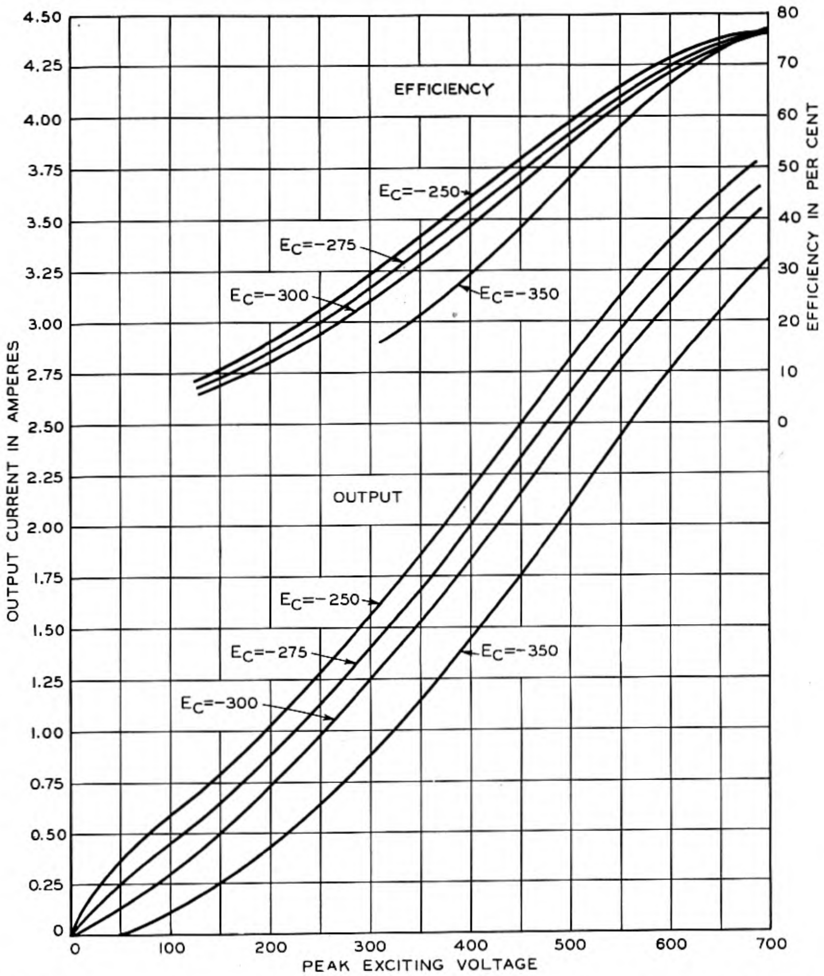


Fig. 13—Dynamic output and efficiency of a three-element tube. $E_b = 3000$ volts, $\mu = 10$, $Z_0 = 1500$ ohms.

screen grid tube which is identical in construction with the three-element tube except for the addition of the screen between the grid and plate. The μ of this tube considering the screen to be the plate is about 4, so that the cut-off bias is about $-E_s/4$. The curves of

Fig. 15 were taken with the screen voltage and output impedance constant at 400 volts and 1500 ohms respectively. It will be noted that these curves show the same general characteristics at the lower portions as the curves of the three-element tube, Fig. 13, but as predicted in the theoretical consideration, they tend to saturate more rapidly

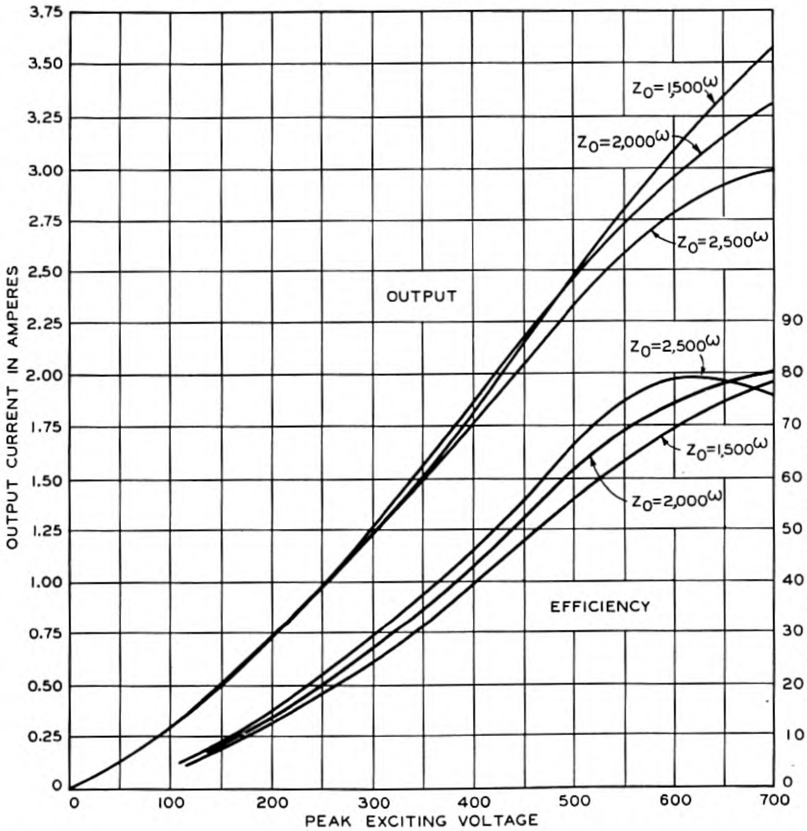


Fig. 14—Dynamic output and efficiency of a three-element tube. $E_b = 3000$ volts, $E_c = -300$ volts, $\mu = 10$.

in the upper portion. The efficiencies appear to be about the same as obtainable from the three-element tube at the same outputs.

Fig. 16 shows a comparison of the dynamic output curves at 1500 ohms impedance and bias at cut-off of the screen grid tube for three different values of screen voltage, and of the three-element tube. In addition the d-c. screen currents and grid currents are shown. The high μ effect of the screen grid tube is responsible for the steeper dynamic curve. The reason for the greater saturation effect at the upper portion of the curves for the screen grid tube is apparent upon com-

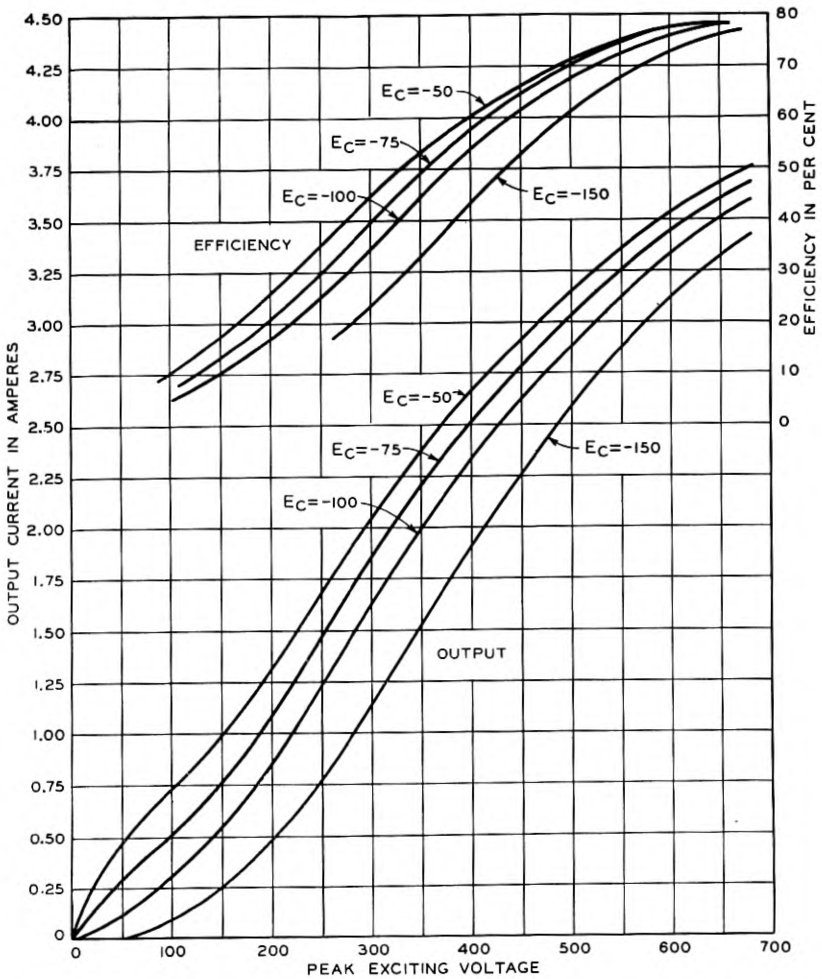


Fig. 15—Dynamic output and efficiency of a screen grid tube. $E_b = 3000$ volts, $E_s = 400$ volts, $Z_0 = 1500$ ohms.

paring the d-c. screen and grid currents with the grid current for the three-element tube, since the plate is being robbed of the current taken by grid and screen in one case, and only of the current taken by the grid in the other. In Fig. 17, the effect of the output impedance on the screen grid dynamic curves is shown. The effect is seen to be similar to that for the three-element tube, though somewhat more pronounced.

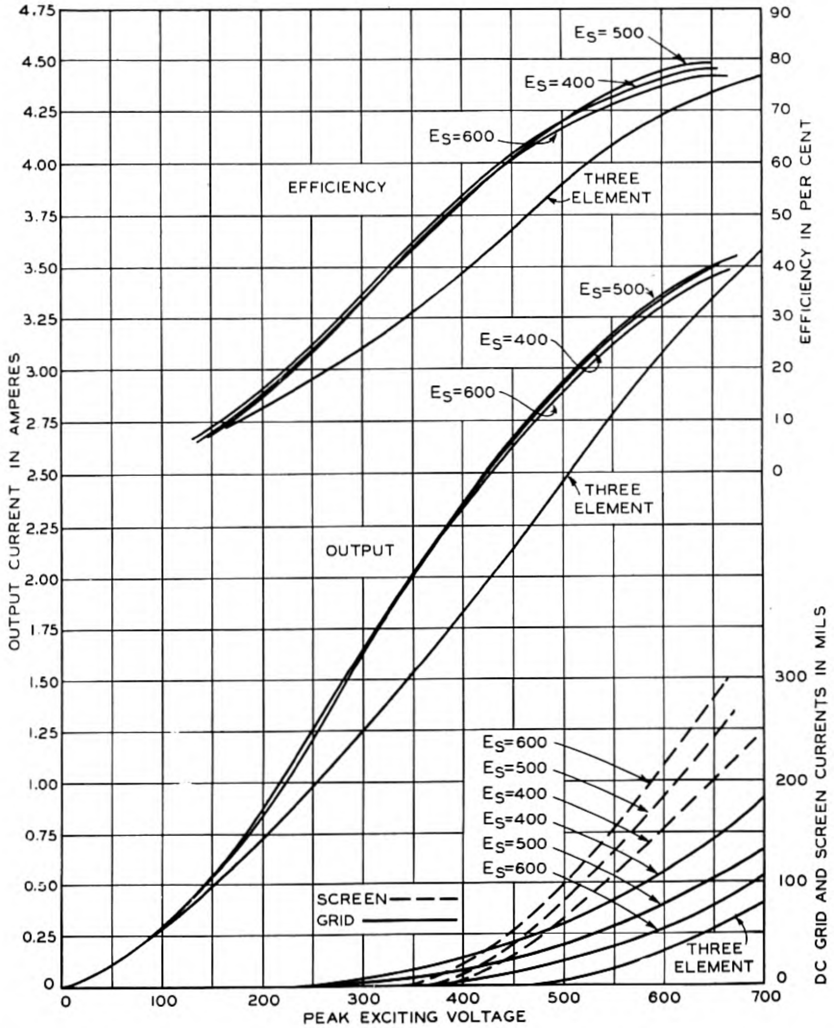


Fig. 16—Dynamic output and efficiency of a screen grid tube compared to that of a three-element tube. $E_b = 3000$ volts, E_c at cut-off $Z_0 = 1500$ ohms.

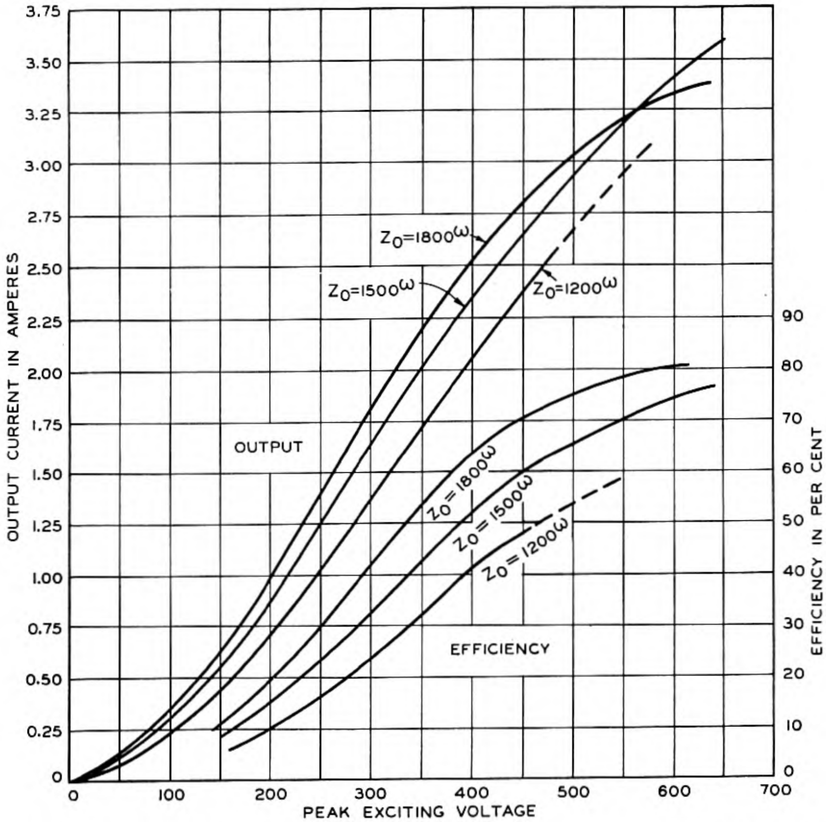


Fig. 17—Dynamic output and efficiency of a screen grid tube. $E_b = 3000$ volts, $E_c = -100$, $E_s = 400$ volts.

Conclusions

It has been shown from both theory and experiment that there is a marked difference between the dynamic output characteristics of Class B and Class C amplifiers as defined, particularly in the lower portion of the curves. This difference is the more pronounced the farther the grid bias voltage is moved from the cut-off point. In general, Class C operation is more efficient than Class B operation because the flow of plate current is limited to a smaller portion of the cycle which includes the portion in which the plate voltage is lowest.

It must be borne in mind that in a radio telephone transmitter where modulation is effected at low power level and the modulated carrier amplified, the dynamic characteristic of the output current will be the resultant of the dynamic characteristics of all of the intermediate stages from the modulating stage through the final amplifier stage. In certain cases distortion in one stage might be fairly well compensated for by suitably shaping the characteristic of the following stage through choosing proper values of bias voltage and output impedance.

It may be concluded from both theory and experiment that the screen grid tube functions in general similarly to the three-element tube and is capable of giving about the same output efficiency. It possesses the property of the high μ tube in giving greater output for a given exciting voltage in the portion of the dynamic characteristic where saturation has not become noticeable, but the dynamic characteristic shows saturation much sooner than that of the three-element tube. The screen grid tube also has the property that the output impedance has very little effect on the input circuit which will allow some shaping of the dynamic characteristic without affecting the driving power.

The fact should be mentioned that the so-called internal impedance, or plate impedance of the tube, such as might be measured on a bridge for small amplitudes of plate current swing, bears no very close relation to the impedance of the output circuit into which it should work in Class B or Class C operation. Reference to this plate impedance is quite misleading unless the part of the characteristic from which it is taken is given, since it varies greatly over the characteristic. For the screen grid tube mentioned the internal plate impedance is of the order of 100,000 ohms, whereas it was working with an output impedance of the order of 1500 ohms.

Table of Symbols

E_b = d-c. plate voltage

E_c = d-c. grid voltage

E_s = screen grid voltage (held constant)

e_p = instantaneous plate potential

e_g = instantaneous grid potential

e_{pm} = minimum plate potential

e_1 = instantaneous value of fundamental frequency component of alternating plate voltage

e_n = instantaneous value of n^{th} harmonic component of alternating plate voltage

I_b = d-c. plate current

I_p = peak value of plate current

i_p = instantaneous value of plate current

i_0 = instantaneous value of alternating component of plate current

i_1 = instantaneous value of fundamental frequency component of alternating plate current

i_n = instantaneous value of n^{th} harmonic component of alternating plate current

I_1 = peak value of fundamental frequency component of alternating plate current

I_n = peak value of n^{th} harmonic component of alternating plate current

μ = amplification factor of tube = $\left[\frac{\Delta e_p}{\Delta e_g}\right]_{i_p \text{ constant}}$

K = I_1/I_p

W_0 = power output in watts

Z_0 = output impedance = R_0, a resistance at fundamental frequency

R_p = instantaneous internal resistance of the tube

r_p = differential plate resistance of tube at any point

$$= \left[\frac{\Delta e_p}{\Delta i_p}\right]_{e_g \text{ constant}}$$

The Time Factor in Telephone Transmission *

By O. B. BLACKWELL

Until comparatively recent years the telephone engineer gave little attention to transmission time in his problems. For all practical purposes he could assume that speech was transmitted instantly between the ends of telephone circuits. The rapid extension of the distances over which commercial telephony is given and the introduction of long telephone cables has changed the situation and has introduced time problems in telephone transmission which are of large technical interest and difficulty. As a result, time problems are receiving more consideration in the technical papers published in recent years on transmission. The accompanying bibliography lists a considerable number of such papers. There seems to be no paper, however, giving a general over-all picture of this subject. The present paper gives briefly such a picture.

THE time factor introduces five different types of problems in telephone transmission:

1. *A Slowing-Down of Telephone Communication.* In talking over long lengths of certain types of cable, the time interval between the formation of a sound by the speaker and its reception by the listener may become of sufficient magnitude to slow down conversation. This is not a serious matter with the types of circuits now used in the United States, even for the longest distances between points in this country. It does, however, become of considerable importance when we consider the joining together of long lengths of cable in this country and long lengths in Europe with possibly long lengths of intervening submarine cable.

2. *Delay Distortion.* Difference in the speed of transmission over a circuit of the different frequencies which make up speech. This may introduce peculiar distortions in speech which cause considerable interference.

3. *Echo Effects.* These arise from the fact that parts of the energy transmitted over a circuit may be reflected back from points of irregularity in it, particularly at the ends. Small amounts of the energy may wander back and forth over a circuit two or more times. While these echoes may affect both the talker and listener, they generally have the greatest effect on the talker who may have an uneasy feeling that the distant party wishes to break in on the conversation.

4. *Effects of Voice-Operated Devices.* To overcome echoes, and

* Presented at the South West District Meeting of the A. I. E. E., Kansas City, Mo., October 22–24, 1931.

under some conditions to hold circuits stable, it has become the practise of connecting into certain types of circuits, relay devices operated by the transmitted speech currents which render inoperative transmission in the opposite direction. In some cases delay in transmission may be an advantage in the operation of such devices. In other cases it may introduce serious difficulties. Conditions may be set up in which it is difficult for one party to interrupt the other. In other cases, portions of conversations may be locked out. If the voice-operated devices are not properly adjusted or if considerable noise is present, the devices may not function properly and speech mutilation may result.

5. *Fading.* In radio, the well-known phenomenon of fading is due to waves arriving at the receiver over different paths, the transmission times of which are such as to cause alternate strengthening and weakening of the received signal by alternate phase agreement and opposition. While this factor is mentioned here for completeness, it will not be discussed further as it is beyond the scope of this paper to discuss the problems introduced when there is more than a single path between the sending and receiving ends of a circuit. The present paper is limited to the conditions which hold where not more than one path is involved in the transmission in each direction.

Speed of Transmission

Before considering these problems in more detail it would be well to define what is meant by the speed of transmission over a circuit. There are several speeds which may have significance according to the problem involved.

Whenever a change in applied voltage is made at one end of a circuit, some evidence of this is transmitted over the circuit to the receiving end at the speed of light. In general, however, except in radio, no sufficient action to be of use is transmitted at this speed and it is largely of theoretical importance.

The speed which the engineer generally has in mind in thinking of line transmission is the speed at which the crests or the troughs of the waves pass along the line when a single-frequency potential is continuously applied at the sending end. This usually is referred to as the speed of phase transmission in the steady state. While this usually approximates the speed in which we are interested, it may in particular cases differ considerably from it. In fact, in certain types of artificial circuits the crests and troughs of the waves travel toward rather than away from the sending end.

This speed may best be explained as follows:

Since the speed of transmission in generally different for different parts of the frequency range, for simplicity a particular narrow frequency range, say between the frequencies N_1 and N_2, is considered. It is supposed that electrical filters are applied to the circuit limiting the frequencies over it to approximately this range. If then, a voltage having a frequency, say at midpoint of this narrow range, is applied to the circuit for a short interval and then removed, the speed at which the disturbance thus set up travels down the circuit is the speed in which we are interested. A spurt of energy of this type is evidently similar to that which takes place in carrier telegraph systems when a dot impulse is applied to the circuit. This speed can be looked at, therefore, as that of carrier telegraph signals so formed.

The speeds on this basis of a number of standard constructions which represent good engineering practise today are approximately as follows:

Type of circuit	Approximate speed in miles per second
Cable circuits loaded with 88-mh. coils at 3,000-ft. spacing	10,000
Cable circuits loaded with 44-mh. coils at 6,000-ft. spacing	20,000
Cable pairs of non-loaded 16 B. & S. gage	130,000
Non-loaded open-wire pairs	180,000
Radio	186,000

Causes of Time Lag in Transmission

A pair of wires of zero resistance in free space separated from all other conductors and without leakage would transmit electrical waves over it at the speed of light. It will be noted from the above table that non-loaded open wires transmit at a speed not differing widely from this. What retardation exists comes largely from the glass insulators which cause an increase in capacity and the resistance of the wires, which causes an effective increase in inductance.

In cable circuits there is still further retardation by the increase in the capacity between the wires because of the necessity of using a certain amount of solid dielectric and particularly from the increase in the inductance of the wires when loading coils are inserted in them to decrease attenuation.

In actual circuits there is still some further retardation by the apparatus which is necessarily inserted at the terminals and at intermediate points along the circuit. The figures given in the above table are for the bare circuits. The delays caused by apparatus will, in general, reduce these speeds from 10 to 25 per cent.

Slowing-Down of Telephone Conversation

Considering the first of the above factors, it is noted that so long as the speaker at one end of a telephone circuit continues to talk, the listener at the other end will hear the speech in proper time relation, independent of how much absolute delay there is in going from one end of the circuit to the other. However, when the speaker asks a question and waits for the answer, the slowing-down effect on his conversation will evidently be the time of transmission of his question to the distant end and the transmission back from the distant end of the answer. Considering the speed shown in the table, however, consideration may be taken of the fact that the non-loaded constructions, both open wire and cable, and the radio, are of such high speed that conversations could be carried on over them for the longest distance between places in the world without appreciable difficulty. This, however, is not the case for the loaded construction. Assume, for example, that a length of 4,000 miles would cover the wire line distance between any two points in this country. For the slowest construction noted, an interval of 0.8 second is required for transmission to the distant end and return. While it is possible to carry on conversation over a circuit with this delay, it is larger than is considered desirable. The faster of the loaded constructions shown would give a delay over circuits of this length which represents somewhere about the limit of what, at the present time, is considered satisfactory. Incidentally, the slowest of the constructions shown, for this and other reasons, is not proposed for use except for comparatively short distances.

Communication engineers must look forward to the time when the longest cable distances in North America are connected, in some cases by submarine cable, to long lengths of cable in Europe. With this as an ultimate objective, this matter of the direct effect of delay on conversation has become of considerable interest.

An appreciation of the transmission time on long telephone circuits may perhaps be gained by considering the distance required to produce an equivalent delay of sound waves traveling in air. For example, it takes about as long for a radio wave to travel half way around the earth at the equator as it does for a sound wave to travel from one speaker to another when the distance separating them is about 75 ft. Incidentally, the time required for a radio wave to travel from the earth to the planet Mars would be from about 3 to 20 minutes, assuming that it got there at all. Evidently if we have neighbors on Mars we can never hope to carry on conversation with them.

Telephone engineers have devoted considerable attention to the effects of delay on the telephone users in an effort to determine how far the electrical waves should be permitted to travel over different types of circuits. Since the present constructions do not offer any particular difficulties, for the distances now in use, a look into the future has been taken by means of artificial delay circuits. One method has been to loop back and forth loaded conductors in a cable until the desired delay was obtained. Another method has been to record the

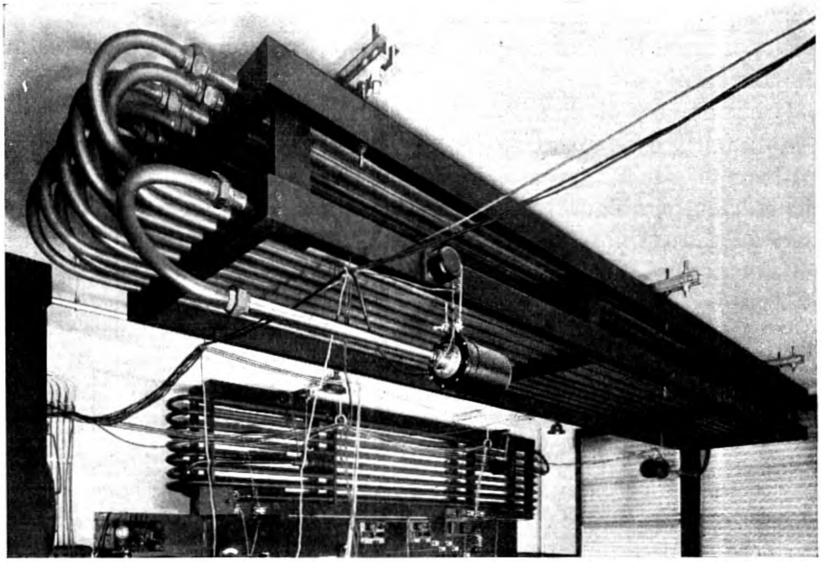


Fig. 1—Acoustic delay circuits.

talkers' waves on a phonograph and pick up the impressions with a second needle, displaced the desired amount from the first so as to introduce delay into the conversation.

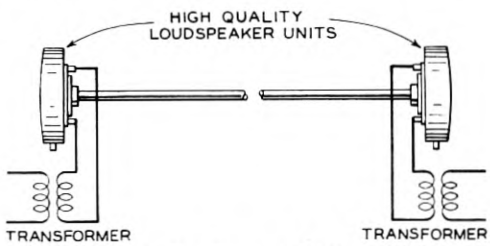


Fig. 2—Acoustical delay circuit showing two high-quality loudspeakers with one connected to each end of tube forming delay circuit.

Considerable use has also been made of pipes or "acoustic" delay circuits. Fig. 1 shows an illustration of a brass pipe delay circuit used in experimental work. Fig. 2 shows the circuit in schematic form. In addition to the pipe, which is looped back and forth to conserve space, the circuit involves telephone receivers at the two ends. The one at the sending end converts electrical energy into sound which is transmitted through the pipe and the one at the receiving end converts the sound waves back into electrical energy. The pipes are quite suitable because they have approximately the same delay at all frequencies. Various devices are required to reduce the reflections which occur at the junction of the pipe and the receiver and to equalize the attenuation.

Using devices of this nature, experimenters have found it possible to talk fairly conveniently over circuits representing time intervals as great as 0.7 second in each direction. So great delays would be considered undesirable for commercial use, however. Delays of about a third of this, in general, are considered about the maximum which is satisfactory.

Delay Distortion

In designing circuits which are electrically long, care must be exercised to insure that the transmission times for all frequencies in the transmission range are sufficiently alike to avoid objectionable transient phenomena. These effects may occur in one-way circuits as well as in two-way circuits and are not related to echo effects.

The appearance of these transients to the listener depends on whether the excess delay is at low frequencies or at high frequencies. It is rather difficult to describe the characteristic sound of a circuit with low-frequency delay. A high-frequency delay, if it is in an extreme form, sounds as though a high-pitched reed, such as a harmonica reed, was being plucked whenever there is a sudden transition in the voice sounds being transmitted over the circuit.

The characteristic effects of transients are conveniently described by the aid of oscillograms of spurts of alternating current taken before and after being sent over circuits having various delay characteristics. To begin with, it must be recalled that when any wave shape is applied to a circuit, the transmitted wave in the circuit can be expressed as the sum of the series of sinusoidal waves whose frequencies range from very low to very high values.

In the case where a sinusoidal wave of frequency F is suddenly applied to the sending end of the line, the effect may therefore be explained as due to an infinity of sinusoidal waves so proportioned and phased as to add up to zero, up to the instant of application of the wave, and to equal the steady-state value of the wave at that instant. Of these waves, the most important have frequencies close to F. They are propagated over the line individually with a velocity corresponding to the frequency. If the velocity of the line is the same for all frequencies, they will evidently add up to give the same over-all wave shape at the receiving end of the line as at the sending end.

If, on the other hand, the velocity is not the same for all frequencies, there will be more or less distortion and transients in establishing the wave, though ultimately the pure wave of frequency F will be established. Several oscillograms will be shown to indicate transient effects which are experienced under various conditions.

Fig. 3 is an oscillogram showing a spurt of 1,600-cycle current as applied to and received from a loaded circuit having fairly large delays in the upper part of the transmitted range compared to the delay at

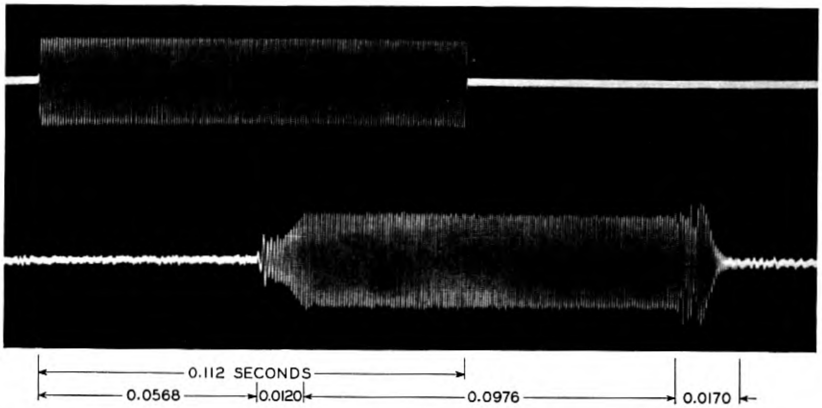


Fig. 3—Transients in 522 miles of medium heavy loaded repeatered circuit. Upper trace—transmitted 1600-cycle wave. Lower trace—received wave.

lower frequencies. Remembering the nature of the oscillations at the beginning and end of the applied spurt, it will be observed that the current at the receiving end consists at first of a fairly low frequency which builds up in frequency and magnitude to the steady-state value. At the end of the spurt the same transient is experienced, but in this case the higher frequency currents which have been delayed in the line are at the tail end of the train.

Fig. 4 shows a 200-cycle current with many harmonics of higher degree transmitted over a circuit having large delay at low frequencies. It will be noted that these high or harmonic frequencies are received in advance of the 200-cycle wave. This is because the 200-cycle wave is subject to appreciable delay while the higher frequencies are not. This circuit, while actually made up of artificial networks, had characteristics similar to certain types of long cable circuits for the lower part of the telephone frequency range.

Fig. 5 shows transient effects of a 600-mile composited 19-gage H-174 side circuit. The zero lines of the three curves of this figure show slight effects of crosstalk and other interference. These effects are not, however, of sufficient magnitude to interfere with the general appearance of the signals. The upper line shows the applied 1,000-cycle current.

Fig. 4—Transients in 174 sections of high-pass filter (cutoff frequency 107 cycles). Upper trace—transmitted 200-cycle wave with harmonics obtained from an overloaded amplifier. Lower trace—received wave.

The next line shows the current as it was received at the end of the line. The transients which are produced at the beginning and ending of the signals are evident.

By the insertion of proper networks it is possible to correct for distortions of this kind. In the last line there is shown the received cur-

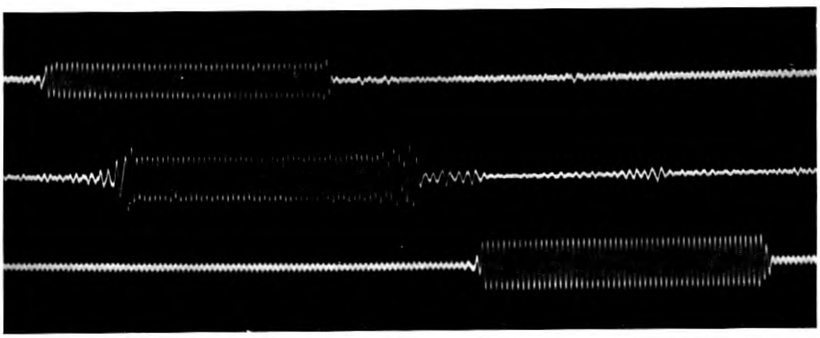


Fig. 5—Correction of transients in 600 miles of medium heavy loaded cable circuit through use of delay correcting network. Upper trace—transmitted 1000-cycle wave. Middle trace—wave received from line and applied to delay correcting network. Lower trace—wave received from delay correcting network.

rent when a delay correcting network of this kind is applied in series with the circuit. It should be noted that while the delay in the reception of the signal is somewhat increased (as shown by the displacement of the whole signal farther to the right) the transients at the beginning and stopping of the signal are very much reduced.

Echoes

In designing telephone circuits which are electrically long, an important problem is presented by the necessity of avoiding echo effects. These are caused by reflection of electrical energy at points of discontinuity in the circuit and are very similar to echoes of sound waves in an auditorium. The reflected waves are usually considered as echoes when there is an appreciable delay with respect to direct transmission. Some of the reflected waves return to the receiver of the talker's telephone so that if the effects are severe he may hear an echo of his own words. Other reflected waves enter the receiver of the listener's telephone and, if severe, cause the listener to hear an echo following the directly received transmission.

Reflections of voice waves occur in all practical telephone circuits. It is only in telephone circuits of such length as to require a number of repeaters, however, that echo effects become serious. The fact that the circuits are electrically long makes the time lag of the echoes appreciable. At the same time, the telephone repeaters overcome the high attenuation in these long circuits and consequently make the echoes louder. The seriousness of the effect is a function of both the time lag and the volume of the echo relative to the direct transmission, becoming greater when these are increased.

In telephone circuits the most important points of discontinuity are usually the two ends of the circuit. In a four-wire telephone circuit these are the only points of discontinuity.

Fig. 6 shows a schematic diagram of a four-wire telephone circuit and a schematic representation of the direct transmission over the circuit, together with the various talker and listener echoes which are set up. The rectangles at the extreme right and left are intended to represent the telephone sets used by two subscribers at the west and east terminals of the circuit. The rectangles marked N represent electrical networks which simulate or balance more or less perfectly the impedance of the telephone sets. In the four-wire circuit the rectangles with arrows represent one-way repeaters or amplifiers. At each terminal the two separate one-way circuits comprising the four-wire circuits are joined together by means of the familiar balanced transformers. When the subscriber at W talks, the transmission passes to E over the upper path in the four-wire circuit. This is indicated by the heavy line labeled "Direct Transmission" in part b of the figure. When subscriber E talks, transmission passes over the lower path in a similar manner.

Considering part b of the figure, it will be noted that when direct transmission is received at the east end of the circuit, a portion of the current passes to the opposite side of the four-wire circuit and is transmitted to the subscriber at the west end as a talker echo. Similarly, a portion of this talker echo is transmitted over the upper part

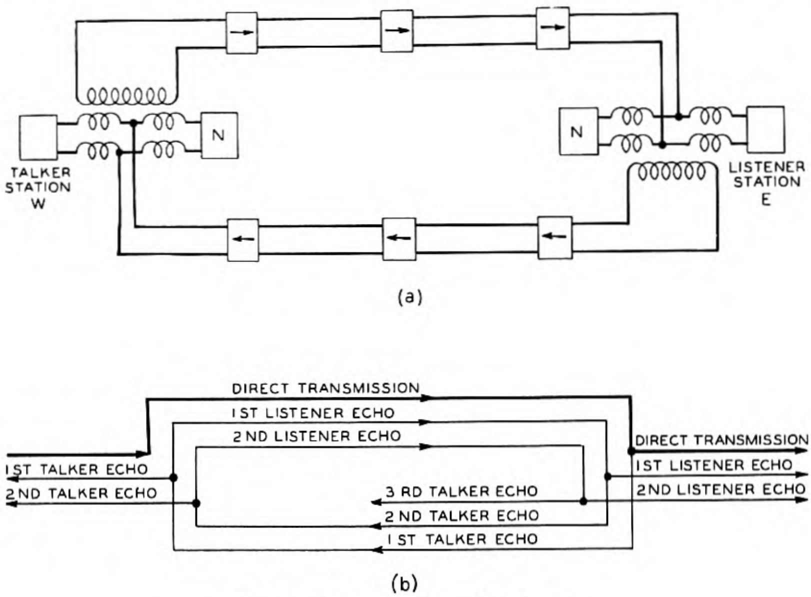


Fig. 6—Echoes in four-wire circuit.

of the circuit to the listener at the east end of the circuit as a listener echo. Successive talker and listener echoes follow this, as indicated in the diagram. If the networks at the two ends of the circuit can be made to simulate accurately the subscriber circuits, none of these echoes will exist. A high degree of simulation, however, is impracticable in an economical telephone plant under usual conditions. In two-wire circuits with many repeaters the echo paths may become very complicated.

An interesting case of "echoes" is that which may be produced when two radio stations are sending out the same program at the same wavelength. The program as received by one of these stations over wire circuits is, of course, slightly delayed with respect to a station nearer the source of the program. It is possible, then, for a receiving set properly located to receive both of these stations, in which case if the time difference is sufficient one of them will sound like an echo of the other. In this case there may be the added peculiarity that if the weaker station is the one at which the program is received first it will appear that the "echo" is in advance of rather than following the sound which appears to cause it.

Time Effects with Voice-Operated Devices

Switching devices operated by the voice currents themselves are frequently introduced into long telephone circuits. In general, the effect of such devices is to render inoperative transmission in the direction opposite to that of the speech waves which are going over a circuit at the particular instant. The first use of any considerable importance to which such devices were put was in connection with long circuits for the purpose of preventing the building-up of undesirable echoes. More recently, however, long radio telephone circuits have come into use. These circuits may vary rapidly in transmission effectiveness. If these circuits are arranged to be operative in both directions at a time it would be very difficult to prevent their becoming unstable and possibly setting up oscillations. For this reason such circuits are frequently operated with switching arrangements such that the circuits leading to both transmitting stations are normally disabled and rendered inoperative. When a subscriber speaks at either end, therefore, the voice currents must operate switching devices which restore the circuits leading to his transmitting station. Incidentally, this must render inoperative the receiving circuit at the same time.

A very interesting application of time delay has been made in connection with radio systems so operated. In this arrangement the voice currents when they reach the disabling point are passed through an artificial line in which a desired amount of delay is incorporated. Just before entering this line a fraction of the energy is taken, rectified, and made to operate the switching mechanism for restoring the circuit to operating condition. This switching is so arranged as to be completed by the time that the voice currents have passed through the artificial delay circuit and are ready to proceed down the line. If it were not for this arrangement a small part of the speech currents might be dissipated during the interval while the switching mechanisms were operating.

Fig. 7 is similar to Fig. 6 noted above with the exception that the application of an echo suppressor is shown.

When the subscriber at the left of the drawing begins to talk, the waves set up in his telephone are transmitted over the upper part of the circuit. Upon reaching the input of the echo suppressor, a small part of the energy is diverted to short circuit the lower branch of the circuit, as indicated. Meanwhile, the main transmission passes on to the subscriber at the right. Echoes which return in the lower part of the circuit are blocked as indicated. After the talker has ceased speaking, the device remains operative for a time equal to the delay

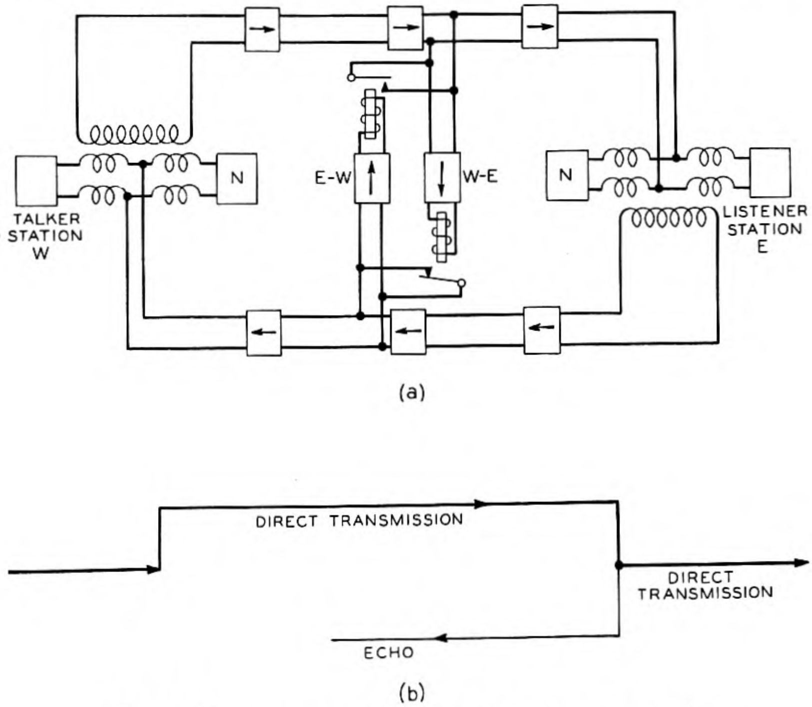


Fig. 7—Echo suppressor cutting off echo in four-wire circuit.

of the echo as measured from the input of the device to the disabling point plus an additional time to take care of echoes in the circuit between the four-wire terminal and the subscriber.

When the suppressor releases, the circuit is again free to transmit in either direction. When the right-hand subscriber talks, the action is similar except that the other half of the echo suppressor operates.

In practical use the echo suppressors are so carefully controlled that telephone users are generally unable to tell whether a suppressor is on the circuit or not. This is due to the short delays and careful adjustments of the time functions of the device. It is possible, however, if the delays are longer and the adjustments are made with less care, to introduce two types of difficulty. If one subscriber talks fairly steadily he may hold the suppressor operative so continuously that it is difficult for the other party to break in on the conversation if he so desires. If one subscriber started to reply almost simultaneously with the termination of the other's speech, part of his reply might be blocked at the echo suppressor along with the last of the echo.

Further difficulties arise if two circuits, each containing an echo suppressor of this kind, are switched together in tandem. In this case it would be possible for the subscribers to completely block each other's speech if they started talking simultaneously.

In the case of radio circuits operated as noted above, difficulties are introduced somewhat similar to that of two echo suppressors in tandem.

The above will sufficiently suggest the types of difficulties which arise from delay in connection with voice-operated devices. Certain of the papers in the accompanying bibliography consider these problems in more detail.

In preparing the accompanying bibliography, no attempt has been made to make it complete. It is believed, however, to contain most of the important publications on the subject.

Bibliography

General

"Telephone Transmission over Long Cable Circuits," A. B. Clark. *Bell System Technical Journal*, vol. 2, no. 1, Jan. 1923, p. 67–94; A. I. E. E. Trans., vol. 42, Feb. 1923, p. 86–96; *Electrical Communication*, vol. 1, no. 3, Feb. 1923, p. 26–40.

"Development and Application of Loading for Telephone Circuits," Thomas Shaw and William Fondiller. *A. I. E. E. Trans.*, vol. 45, Feb. 1926, p. 268–292; *Bell System Technical Journal*, vol. 5, no. 2, April 1926, p. 221–281; *Electrical Communication*, vol. 4, no. 4, April 1926, p. 258–276, and vol. 5, no. 1, July 1926, p. 38–53.

"Extending the Range of Loaded Lines by Echo Suppression and Phase Equalization," K. Küpfmüller. *Elektrische Nachrichten-Technik*, vol. 3, no. 3, Mar. 1926, p. 82–88.

"Transmission Characteristics of Long Distance Telephone Cable Connections," K. Küpfmüller. *Europäischer Fernsprechdienst*, no. 5, July 1927, p. 19–33.

"Telephone Communication System of the United States," Bancroft Gherardi and F. B. Jewett. *Bell System Technical Journal*, vol. 9, no. 1, Jan. 1930, p. 1–100.

"Recent Developments in Toll Telephone Service," W. H. Harrison. *A. I. E. E. Trans.*, vol. 49, no. 1, Jan. 1930, p. 166–176.

"Some Recent Developments in Long Distance Cables in the United States of America," A. B. Clark. *Bell System Technical Journal*, vol. 9, no. 3, July 1930, p. 487–492.

"Transmission Characteristics of Open-Wire Telephone Lines," E. I. Green. *Bell System Technical Journal*, vol. 9, no. 4, Oct. 1930, p. 730–759.

"Range and Economy of the Toll Cable Lines of System II of the C. C. I.," K. Höpfner and F. Lüschen. *Europäischer Fernsprechdienst*, no. 21, Jan. 1931, p. 3–16.

Delay Distortion

"Theory of the Transient Oscillations of Electrical Networks and Transmission Systems," J. R. Carson. *A. I. E. E. Trans.*, vol. 38, pt. 1, 1919, p. 345–427.

"Transient Phenomena in Wave Filters," K. Küpfmüller. *Elektrische Nachrichten-Technik*, vol. 1, no. 5, Nov. 1924, p. 141–152.

"Measurement of Phase Distortion," Harry Nyquist and Smart Brand. *Bell System Technical Journal*, vol. 9, no. 3, July 1930, p. 522–549.

"Building-Up of Sinusoidal Currents in Long Periodically Loaded Lines," J. R. Carson. *Bell System Technical Journal*, vol. 3, no. 4, Oct. 1924, p. 558–566.

"Attenuation of Band-Pass Filters in the Transmission Range," H. F. Mayer. *Elektrische Nachrichten-Technik*, vol. 2, no. 10, Oct. 1925, p. 335–338.

"Phase Distortion and Phase Distortion Correction," S. P. Mead. *Bell System Technical Journal*, vol. 7, no. 2, April 1928, p. 195–224.

"Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks," O. J. Zobel. *Bell System Technical Journal*, vol. 7, no. 3, July 1928, p. 438–534.

"Effects of Phase Distortion on Telephone Quality," J. C. Steinberg. *Bell System Technical Journal*, vol. 9, no. 3, July 1930, p. 550–566.

"Phase Distortion in Telephone Apparatus," C. E. Lane. *Bell System Technical Journal*, vol. 9, no. 3, July 1930, p. 493–521.

Echoes

"Irregularities in Loaded Telephone Circuits," George Crisson. *Bell System Technical Journal*, vol. 4, no. 4, Oct. 1925, p. 561–585; *Electrical Communication*, vol. 4, no. 2, Oct. 1925, p. 98–110.

"Limitation of the Gain of Two-Way Telephone Repeaters by Impedance Irregularities," George Crisson. *Bell System Technical Journal*, vol. 4, no. 1, Jan. 1925, p. 15–25.

"Connecting-Together of Four-Wire Lines," K. Höpfner. *Telegraphen- und Fernsprech-Technik*, vol. 16, no. 6, June 1927, p. 166–169.

"Measurement of Echo Attenuation," R. Feldtkeller and H. Jacoby. *Telegraphen- und Fernsprech-Technik*, vol. 17, no. 3, March 1928, p. 61–71.

Voice-Operated Devices

"Echo Suppressors for Long Telephone Circuits," A. B. Clark and R. C. Mathes. *A. I. E. E. Trans.*, vol. 44, April 1925, p. 481–489; *Electrical Communication*, vol. 4, no. 1, July 1925, p. 40–50.

"Echo Suppressors for Telephone Lines," H. F. Mayer and H. Nottebrock. *Telegraphen- und Fernsprech-Technik*, vol. 15, no. 11, Nov. 1926, p. 353–355.

"New York-London Telephone Circuit," S. B. Wright and H. C. Silent. *Bell System Technical Journal*, vol. 6, no. 4, Oct. 1927, p. 736–749.

"Transatlantic Telephony. The Post Office Differential Voice-Operated Anti-Singing Equipment," C. A. Beer and G. T. Evans. *P. O. E. E. Journal*, vol. 20, no. 1, April 1927, p. 65–72.

"Bridge for Measuring Small Time Intervals," Joseph Herman. *Bell System Technical Journal*, vol. 7, no. 2, April 1928, p. 343–349.

"New Standard Echo Suppressor," W. F. Marriage, P. R. Thomas, and K. G. Hodgson. *Electrical Communication*, vol. 9, no. 3, Jan. 1931, p. 196–202.

Related Transmission

"Transmission of Pictures over Telephone Lines," H. E. Ives, J. W. Horton, R. D. Parker, and A. B. Clark. *Bell System Technical Journal*, vol. 4, no. 2, April 1925, p. 187–214.

"Some Studies in Radio Broadcast Transmission," Ralph Bown, DeL. K. Martin, and R. K. Potter. *Bell System Technical Journal*, vol. 5, no. 1, Jan. 1926, p. 143–213.

"Wire Transmission System for Television," D. K. Gannett and E. I. Green. *Bell System Technical Journal*, vol. 6, no. 4, Oct. 1927, p. 616–632.

"Some Developments in Common Frequency Broadcasting," G. D. Gillett. *Proceedings of the Institute of Radio Engineers*, vol. 19, no. 8, Aug. 1931, p. 1347–1369.

Constant Frequency Oscillators *

By F. B. LLEWELLYN

Summary—The manner in which the frequency of vacuum tube oscillators depends upon the operating voltages is discussed. The theory of the dependence is derived and is shown to indicate methods of causing the frequency to be independent of the operating voltages. These methods are applied in detail to the more commonly used oscillator circuits.

Experimental data are cited which show the degree of frequency stability which may be expected as a result of application of the methods outlined in the theory, and also show that the best adjustment is in substantial agreement with that predicted by theory. With a carefully built and adjusted oscillator the effects of normal variations in the operating voltages are negligible in comparison with the effects of temperature variations resulting from the changed operating currents. Methods of preventing these latter effects are not discussed in the present paper.

The appendix contains an analysis of the conditions under which the performance of an oscillator may be represented by the use of linear circuit equations.

IN recent years the commercial requirements of vacuum tube oscillators have grown more rigid. The tremendous increase in the number of radio broadcast stations with the consequent narrowing of frequency band available to each, the analogous demands by the carrier telephone, and the tendency toward higher frequencies where a small percentage frequency change defeats the universal effort to secure better quality, all have united in creating a need for very constant frequencies. This need has led to a study of methods for holding the frequency constant. The most notable of these is the piezo-electric crystal. However, it has been known for some time that certain oscillator circuits have the inherent property of maintaining their frequency quite constant even though not crystal controlled. Some of these circuits have the additional advantage of combining constant frequency at a given wave-length with the ability to maintain this constancy at other wave-lengths, thus giving a range of available frequencies, any one of which may be depended upon to stay constant.

The elements which cause the frequency of oscillators which are not crystal controlled to vary are such things as vibration, changing temperature, fluctuating voltage, and changing load. Vibration and temperature affect primarily the inductance and capacity in the circuit which naturally causes the frequency to change. Fluctuating voltages change the tube resistance, which in turn affects the frequency.

* Presented at Sixth Annual Convention of I. R. E., Chicago, Illinois, June 4–6, 1931. Published in *Proc. I. R. E.*, Dec., 1931.

Changing loads also change the frequency, since they take the form of variable resistance and reactance.

Considerable work has been done by various individuals to make the inductance and capacity of standard apparatus as free from the effects of vibration and temperature changes as possible. Work of that kind is not discussed here. Variable voltages occur in practically all installations and changing load impedance in many. These two things are the actual cause of the larger part of frequency variation in many installations.

Several vacuum tube circuits have been devised to surmount these difficulties. They may be divided roughly into two groups: first, those in which the attempt has been to minimize the change of frequency with battery voltage, and second, those in which the attempt has been to prevent the change. In the first group we have two types: first, circuits in which the effective resistance has been reduced to as low a figure as possible, and second, those in which a high impedance has been inserted between the tube and the tuned circuit in order to reduce the relative effect of changes in the tube. Considerable success has attended the efforts of a number of engineers led by J. W. Horton in these directions. More recently, circuits of the first type have been applied to the production of relatively high frequencies.[1]

The second group in which the attempt has been to prevent the frequency change developed from the work of Messrs. J. F. Farrington and C. F. P. Rose. They found that a certain critical value of an impedance between the vacuum tube and the tuned circuit apparently produced a constant frequency over a limited range when the battery voltages were varied. They experimented with various forms of networks for this stabilizing impedance and developed several in which the output power was not reduced by stabilization.

Theory

The writer attacked the problem from a theoretical standpoint and showed that in certain cases the mathematical procedure indicates means of making the oscillator frequency independent not only of a variable load resistance, but also of the battery voltages. The purpose of this paper is to develop the general theory and application of these circuits and to show how several circuits in particular may be made to produce practically constant frequency with customary variations of voltage and load resistance. The relations necessary to maintain the frequency constant at any given setting when it is desirable

[1] Ross Gunn, "A new frequency stabilized oscillator system," *Proc. I. R. E.*, **18,** September, 1930.

that the oscillator be operative over a range of frequencies are also indicated.

Before proceeding with a detailed description of the various specific embodiments necessary to secure independence of frequency and battery voltage, it will be well to lay down the physical conditions upon which the frequency of any vacuum tube oscillator depends.

In general, all such oscillators consist of or may be resolved into, a tuned electrical circuit or network to which is attached a vacuum tube. Irrespective of any particular circuit, the frequency of the oscillator is completely determined by the following quantities, the designations used here being uniformly employed throughout the subsequent analysis:

L, the self-inductance in the network
M, the mutual inductance in the network
C, the capacity in the network
R, the resistance in the network
r_p, the plate resistance of the vacuum tube
r_g, the grid resistance of the vacuum tube
μ, the amplification factor of the vacuum tube

Of these quantities, L, C, and M require little comment. They are merely symbolic of the elements of the electrical network. The quantity C includes the interelectrode capacities of the tube when they become of consequence. These three quantities are assumed to be constant, an assumption which has been found very reasonable in practice. The quantity R represents the resistance in the network. For the purpose of this discussion the oscillator is assumed to deliver only a small amount of power, being used most often in such a manner as to supply voltage to the grid of an amplifier tube. Consequently, the electrical network external to the vacuum tube may, and should, be constructed in such a manner as to include a minimum amount of resistance. Under these conditions the losses in the circuit have been found to be practically all the result of the internal resistances, r_p and r_g of the vacuum tube.

These two quantities, r_p and r_g, are very important, being principally responsible for changes in condition of the circuit as a whole. It should be realized that r_g has the same relation to the static values of grid current and potential that r_p has to the plate current and potential. The effect of varying the applied potential of the grid or plate, or of changing the filament current is directly to cause r_p and r_g to vary, usually in opposite directions. Further, when amplitude of oscillation varies, for which variation of battery voltages (grid, plate, and filament) are again principally responsible, both r_g and r_p vary.[2]

The quantity μ is the amplification factor and is used with its usual significance. It varies with battery potential but this variation is ordinarily very small, though not to be neglected.

It eventuates, from the above considerations, that if the reactive elements of the frequency determining circuit are constant, a permissible assumption, the frequency may be stabilized if adequate account is taken of changes in battery voltages and load resistances. This it is the purpose of the present paper to discuss.

Hartley Oscillator

Consider first the form of the Hartley oscillator shown in schematic form without indicating any special method of introducing the batteries, in Fig. 2. Figure 1 shows the circuit equivalent of several of the oscillators in the following figures when the impedances are represented in generalized form, and therefore will be employed for an analysis of the conditions necessary to secure independence of frequency and battery or applied voltages, and the results of this analysis will then be interpreted in detail in terms of the special circuit of Fig. 2. In Fig. 1 the impedances, Z_4 and Z_5, are inserted for the purpose of effecting the independence of frequency and battery voltages, and the values which they should have in order to accomplish this result are found by the following analysis:

From Fig. 1 we have the circuit equations when the assumed current conditions are as shown by the arrows:

$$\begin{aligned}
\mu e &= I_1(r_p + Z_1 + Z_5) + I_2(Z_1 + Z_m) - I_3 Z_m, \\
0 &= I_1(Z_1 + Z_m) + I_2 Z_0 - I_3(Z_2 + Z_m), \\
0 &= -I_1 Z_m \qquad - I_2(Z_2 + Z_m) + I_3(r_g + Z_2 + Z_4), \\
e &= I_3 r_g.
\end{aligned} \tag{1}$$

These equations are expressions of Kirchkoff's Law regarding the sum of the potentials in a closed mesh. The equations (1) effectively comprise only three simultaneous equations because the network has only three meshes.

In the above equation Z_0 is symbolic of the series impedance of the tuned circuit. Using the symbolism of Fig. 1,

$$Z_0 = Z_1 + Z_2 + Z_c + 2Z_m. \tag{2}$$

[2] The appendix to this paper contains a further discussion of the significance of r_p and r_g together with an analysis of the conditions under which oscillator networks may be treated by the use of linear circuit equations as is done in the following analysis.

Equations (1) may be rewritten in determinant form as follows:

$$\begin{vmatrix} (r_p + Z_1 + Z_5) & +(Z_1 + Z_m) & -(Z_m + \mu r_g) \\ +(Z_1 + Z_m) & Z_0 & -(Z_2 + Z_m) \\ -Z_m & -(Z_2 + Z_m) & (r_g + Z_2 + Z_4) \end{vmatrix} = 0. \quad (3)$$

This determinant form of (1) follows immediately from reducing (1) to three equations.

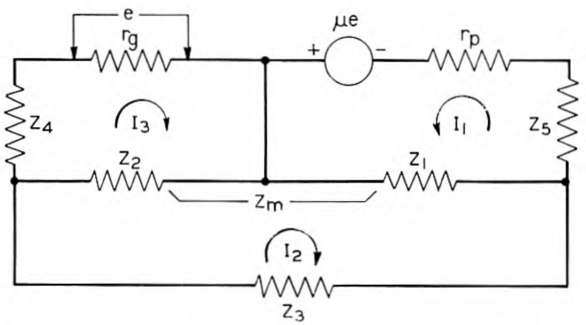


Fig. 1—Equivalent circuit network of Hartley or Colpitts-type oscillator.

In accordance with the theory of the operation of oscillators, discussed in the appendix, both the conditions necessary for oscillation to exist and the frequency of oscillation may be found from (3). That is:

$$\begin{aligned}(r_p + Z_1 + Z_5)Z_0(r_g + Z_2 + Z_4) + (Z_1 + Z_m)(Z_2 + Z_m)(\mu r_g + 2Z_m) \\ = Z_0Z_m(\mu r_g + Z_m) + (Z_1 + Z_m)^2(r_g + Z_2 + Z_4) + (Z_2 + Z_m)^2 \\ (r_p + Z_1 + Z_5). \quad (4)\end{aligned}$$

The next step is to express each of the generalized Z's in the equivalent form of $(R + iX)$ where i stands for the imaginary quantity, $\sqrt{-1}$, and both R and X are real, representing, respectively, resistance and reactance. A great simplification results when it is recalled that the circuits external to the vacuum tube are assumed to have very little resistance, and that practically all of the losses in the network are caused by the tube resistances, r_g and r_p, so that these two are the only resistances which need be retained in the analysis. With this understanding, (4) becomes:

$$\begin{aligned}[r_p+i(X_1+X_5)]iX_0[r_g+i(X_2+X_4)]-(X_1+X_m)(X_2+X_m)(\mu r_g+2iX_m) \\ = -X_0X_m(\mu r_g+iX_m)-(X_1+X_m)^2[r_g+i(X_2+X_4)] \\ -(X_2+X_m)^2[r_p+i(X_1+X_5)]. \quad (5)\end{aligned}$$

In order for (5) to be true, both the real and the imaginary portions must separately be equal to zero. If (5) (which comes naturally from (3) with the given substitutions) is separated into its real and imaginary parts, the resulting two equations must be simultaneous and between them express the frequency and relative values which r_p and r_g must assume in order for oscillations to exist. The particular aim in the present case is to find whether values of X_4 or X_5 exist which will enable the frequency to be expressed in terms of the constants of the circuit external to the vacuum tube so that if r_p, r_g, and μ should vary, the frequency, being dependent upon the external circuit only, will remain constant.

From (5), then, the real and imaginary parts give the following two equations:

$$\begin{aligned} &- X_0[r_p(X_2 + X_4) + r_g(X_1 + X_5)] - \mu r_g(X_1 + X_m)(X_2 + X_m) \\ &\quad = - X_0 X_m \mu r_g - (X_1 + X_m)^2 r_g - (X_2 + X_m)^2 r_p. \qquad (6) \\ &X_0[r_p r_g - (X_1 + X_5)(X_2 + X_4)] - 2X_m(X_1 + X_m)(X_2 + X_m) \\ &\quad = - X_0 X_m^2 - (X_1 + X_m)^2(X_2 + X_4) - (X_2 + X_m)^2(X_1 + X_5). \qquad (7) \end{aligned}$$

There are certain mathematical rules for finding whether the desired constancy of frequency may be obtained from the conditions given by (6) and (7). Without, however, going into detail in regard to these, it is easy to see from (7) that if X_4 and X_5 have such values as to satisfy the equation:

$$\begin{aligned} 2X_m(X_1 + X_m)(X_2 + X_m) &= (X_1 + X_m)^2(X_2 + X_4) \\ &\quad + (X_2 + X_m)^2(X_1 + X_5) \qquad (8) \end{aligned}$$

(which is obtained by including all terms of (7) which do not contain X_0), then the frequency of oscillation is exactly that which will cause X_0 to become zero, and will remain so, no matter what values may be taken by r_p, r_g, and μ. In other words, the oscillation frequency is equal to the series resonant frequency of the tuned circuit.

It follows, then, that if the battery voltages were to vary, the frequency, being determined by the circuit elements external to the vacuum tube only, would remain constant. In regard to a changing load resistance, it is evident that if this were connected in parallel either with r_p or r_g, then the combination of the two resistances could be considered as a single resistance. It therefore follows that the same adjustment which causes the frequency to be independent of battery voltage is also the correct one to render the frequency independent of a variable load impedance when this impedance is resistive, only, and is connected in parallel either with the plate or grid resistance of the tube.

In order to complete the general demonstration, it remains to show that the values imposed on (7) by the condition of (8) do not require physically impossible values of r_p, r_g, and μ in order to satisfy (6) and thus maintain oscillation. To do this, assume that (8) is solved for either X_4 or X_5 and substitute in (6), remembering that X_0 is zero. The result is:

$$\frac{r_p}{r_g} = \mu \left(\frac{X_1 + X_m}{X_2 + X_m} \right) - \left(\frac{X_1 + X_m}{X_2 + X_m} \right)^2. \tag{9}$$

Inspection of this expression shows that the conditions required are physically possible, and it follows that the amplitude of oscillation increases or decreases until the effective values of r_p and of r_g, which are measures of the dissipation of energy on the plate and on the grid sides, take up the values specified by the conditions of (9). Thus, for instance, if X_1 and X_2 were approximately equal, then r_p would have to be $(\mu - 1)$ times as large as r_g before the oscillation amplitude settled down to a steady state value. To many who are accustomed to neglect the losses occurring on the grid side of a vacuum tube when dealing with oscillator problems, this low value of r_g will appear as somewhat unusual. In this connection, it may be pointed out that the low value of r_g is not in any way a special requirement imposed by the stabilizing reactances, X_4 and X_5, but is inherent in vacuum tube oscillators in general, unless particular conditions are arranged to render it otherwise. For instance, it is a well-known experimental fact that resistances of the order of 4000 ohms may be placed across the grid-filament terminals of an oscillator employing any of the more common types of three-element receiving amplifier tubes without stopping the oscillations, when a good low-loss tuned circuit is employed. In view of the fact that the amplitude of the oscillations is commonly limited by r_g, this is evidence that stable oscillations may be secured with values of r_g which are of the order of 2000 or 3000 ohms.

The demonstration may be made more rigid by the use of (6) and for the special case where $X_1 = X_2$ and $X_4 = X_5 = X_M = 0$, in which the stabilizing reactances have been omitted. For such a simplified circuit, it is found by elimination of X_0 between (6) and (7) that:

$$\frac{r_p}{r_g} = \mu \left[\frac{r_p r_g - X_1^2}{r_p r_g + X_1^2} \right] - 1.$$

Now, X_1 is of the order of 500 or 600 ohms at the most, while both r_p and r_g are at least enough larger than this in the case of the more commonly used vacuum tubes so that the expression for r_p/r_g is roughly equal to $(\mu - 1)$. Thus, in the simplest kind of vacuum tube circuit, it is seen that r_g is apt to be appreciably smaller than r_p, and by no means negligible in its effect.

To return to (8), which expresses the relation between X_4, X_5 and other circuit reactances which are necessary to cause the frequency to be independent of battery voltages, we note that, although (8) is still in generalized form, and is yet to be applied to the particular case shown in Fig. 2, the very significant fact that the oscillation frequency for this type of stability must be the series resonant frequency of the tuned circuit is a direct consequence of the requirements of the equation.

For application to the Hartley type of oscillator, the various terms of (8) have the following significance:

$$X_1 = \omega L_1,$$
$$X_2 = \omega L_2,$$
$$X_m = \omega M,$$

where $\omega = 2\pi \times$ frequency and X_4 or X_5 are to be determined. In the case of Fig. 2, where stabilization is accomplished on the plate side, we put X_4 equal to zero. Then solving (8) for X_5, we find:

$$X_5 = 2\omega M\left(\frac{L_1 + M}{L_2 + M}\right) - \omega L_2\left(\frac{L_1 + M}{L_2 + M}\right)^2 - \omega L_1.$$

X_5 is thus required to be negative, so that a capacitive reactance is necessary for plate stabilization of a Hartley-type oscillator. Thus putting

$$X_5 = -\frac{1}{\omega C_5}$$

and remembering that since $X_0 = 0$, the angular frequency is given by

$$\omega^2 = 1/C_3(L_1 + L_2 + 2M),$$

finally we get

$$C_5 = C_3 \frac{L_1 + L_2 + 2M}{L_1 + L_2\left(\dfrac{L_1 + M}{L_2 + M}\right)^2 - 2M\left(\dfrac{L_1 + M}{L_2 + M}\right)}, \qquad (10)$$

which is the value of capacity which should be inserted between the plate and the tuned circuit of a Hartley-type oscillator in order to cause the frequency to remain constant when the battery voltages are varied, and there is no reactance between the grid and tuned circuit.

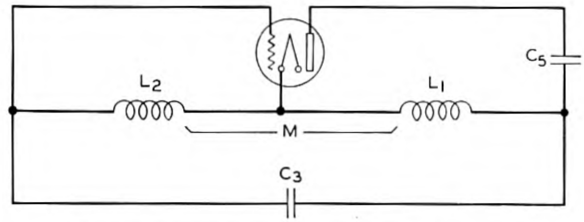


Fig. 2—Hartley oscillator, plate stabilization.

$$C_5 = C_3\left[\frac{L_0}{L_1 + L_2A^2 - 2MA}\right],$$

where

$$L_0 = L_1 + L_2 + 2M, \qquad A = \frac{L_1 + M}{L_2 + M}.$$

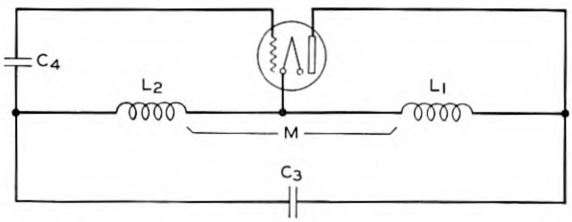


Fig. 3—Hartley oscillator, Grid stabilization.

$$C_4 = C_3A^2\left[\frac{L_0}{L_1 + L_2A^2 - 2MA}\right],$$

where

$$L_0 = L_1 + L_2 + 2M, \qquad A = \frac{L_1 + M}{L_2 + M}.$$

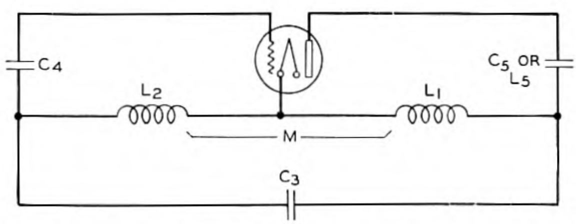


Fig. 4—Hartley oscillator, plate and grid stabilization.

$$\frac{1}{C_5} + \frac{A^2}{C_4} = \frac{1}{C_3}\left[\frac{L_1 + L_2A^2 - 2MA}{L_0}\right],$$

$$L_5 = L_0\frac{C_3}{C_4}A^2 - L_1 - L_2A^2 + 2MA,$$

where

$$L_0 = L_1 + L_2 + 2M, \qquad A = \frac{L_1 + M}{L_2 + M}$$

Of course in actual practice, it is necessary to provide a d-c path for the space current of the tube, which can be accomplished by shunting the condenser, C_5, with a high impedance choke.

It is often the case that a stopping condenser is desirable in the X_4 position, instead of a direct connection between grid and tuned circuit. This stopping condenser and the accompanying leak are advantageous inasmuch as it has been found by experience that an oscillator operating with a leak and condenser combination is inherently much more stable as regards change of frequency with change of battery voltage than an oscillator with a d-c low resistance path from grid to filament, even when a battery is employed to impose a negative bias on the grid. The explanation for this improved stability lies in the fact that the grid leak tends to keep the grid resistance, r_g, constant. It frequently happens, when the leak and condenser combination is used, that difficulty is experienced in avoiding "blocking" when a large enough condenser to have negligible reactance is employed. In such cases the required value of C_5 may be chosen in the manner indicated in connection with Fig. 4, which allows for a finite reactance between grid and tuned circuit, or else, as another alternative, the plate may be directly connected to the tuned circuit so that X_5 is zero, and stabilization may be accomplished by choosing the value of C_4 in accordance with the requirements then imposed by (8), which refer to Fig. 3 and necessitate the value of capacity shown in the figure. Another possible stabilizing arrangement is shown in Fig. 4, where either capacity or inductance may be used on the plate side, depending on the value of capacity at C_4 on the grid side. Yet another possible modification of Fig. 4 would be to use an inductance on the grid side. This would require a very small capacity on the plate side, and probably is less convenient than the arrangement indicated in the figure.

In all three of the cases considered thus far, the equations show that the value of the stabilizing capacity or inductance depends upon the values of L_1, L_2, M, and C_3 so that if the frequency of the oscillator were varied intentionally, by changing L_1, for instance, then a different value of stabilizing capacity or inductance would be required to secure independence of frequency and battery voltage at the new frequency. If, however, the circuit were so constructed that M were zero, and L_1 and L_2 were made so that they remained always equal to each other, then the value of the stabilizing element would depend upon C_3 only, and the frequency could be changed by varying L_1 and L_2 simultaneously without destroying the stabilizing adjustment.

Colpitts Oscillator

This property may be utilized to even greater advantage in connection with the Colpitts type of oscillator, which is illustrated in Figs. 5, 6, and 7 and will now be investigated with the aid of (8) in the same manner in which the relations necessary for stabilizing the Hartley oscillator were secured. Thus, for the Colpitts circuit:

$$X_1 = -\frac{1}{\omega C_1},$$

$$X_2 = -\frac{1}{\omega C_2}.$$

$$X_m = 0,$$

$$\omega^2 = \frac{1}{L_3}\left(\frac{1}{C_1} + \frac{1}{C_2}\right).$$

The values of the stabilizing elements required by (8) are given in Figs. 5, 6, and 7 which show several arrangements for the stabilizing impedance, as applied to the Colpitts-type oscillator. In particular, Fig. 7 shows a choice of either an inductance or a capacity on the grid side.

In Figs. 5 and 6 and in Fig. 7 when inductance is used on both plate and grid sides it is evident that if the condensers, C_1 and C_2, are connected together in a "gang" mounting so that when they are varied, the ratio of their capacities remains constant, then the frequency of the oscillator may be changed by changing C_1 and C_2 without disturbing the stabilizing adjustment which causes the frequency to be independent of battery voltages.

Feed-Back Oscillator

Figures 8, 9, and 10 show conventional drawings of the type of oscillator circuit known as a "feed-back" or sometimes as a "tuned input" circuit. In Fig. 8 stabilizing is accomplished on the plate side; in Fig. 9 on the grid side; and in Fig. 10 on both sides. A mathematical analysis analogous to that which was described in detail in connection with Fig. 1 gives the values of stabilizing impedances which are shown in the figures, and also indicates that the conditions for oscillation may be met when these values are employed.

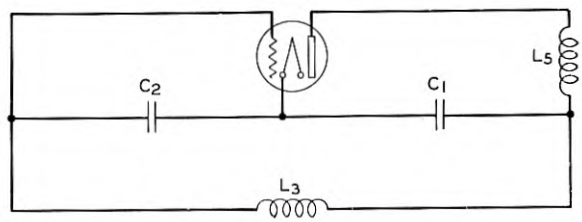


Fig. 5—Colpitts oscillator, plate stabilization.

$$L_5 = L_3 \frac{C_2}{C_1}.$$

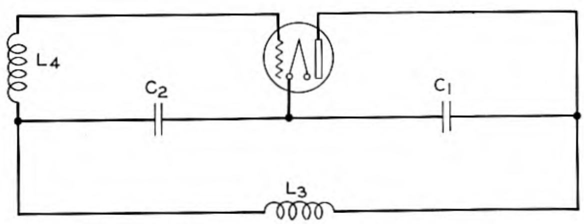


Fig. 6—Colpitts oscillator, grid stabilization.

$$L_4 = L_3 \frac{C_1}{C_2}.$$

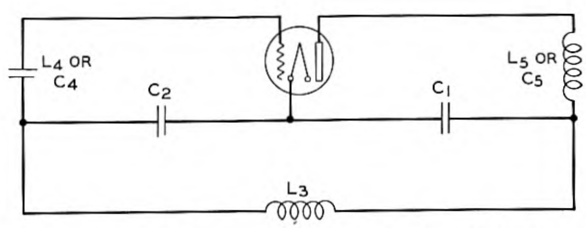


Fig. 7—Colpitts oscillator, plate and grid stabilization.

$$L_4\left(\frac{C_2}{C_1}\right) + L_5\left(\frac{C_1}{C_2}\right) = L_3,$$

$$L_5 = L_3 \frac{C_2}{C_1}\left[1 + \frac{C_2}{C_4}\left(\frac{C_2}{C_1 + C_2}\right)\right].$$

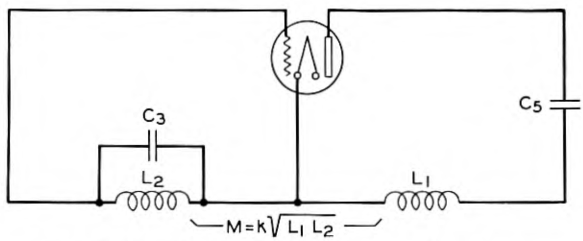


Fig. 8—Feed-back oscillator, plate stabilization.

$$C_5 = C_3 \frac{L_2}{L_1}\left(\frac{1}{1-k^2}\right).$$

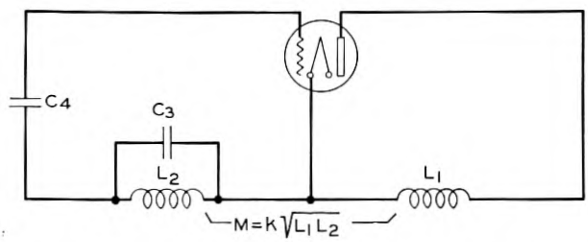


Fig. 9—Feed-back oscillator, grid stabilization.

$$C_4 = C_3\left(\frac{k^2}{1-k^2}\right).$$

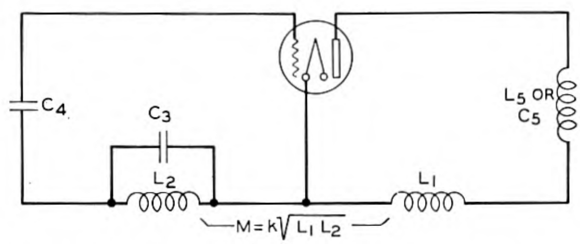


Fig. 10—Feed-back oscillator, plate and grid stabilization.

$$L_5 = L_1\left[k^2\left(1+\frac{C_3}{C_4}\right)-1\right],$$

$$C_5 = C_3\frac{L_2}{L_1}\left[\frac{1}{1-k^2\left(1+\dfrac{C_3}{C_4}\right)}\right].$$

Reversed Feed-Back Oscillator

Figures 11, 12, and 13 show conventional drawings of the type of oscillator circuit known as a "reversed feed-back" or sometimes as a "tuned output" type of oscillator, with the application of stabilizing

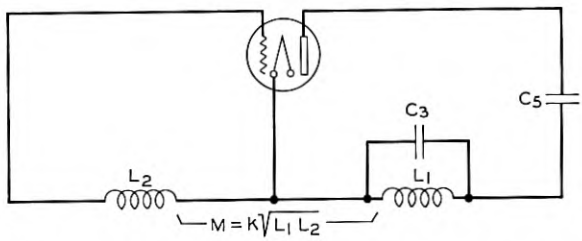


Fig. 11—Reversed feed-back oscillator, plate stabilization.

$$C_5 = C_3\left(\frac{k^2}{1 - k^2}\right).$$

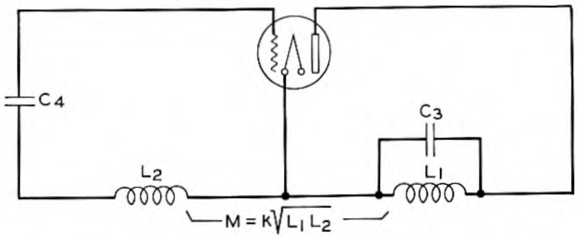


Fig. 12—Reversed feed-back oscillator, grid stabilization.

$$C_4 = C_3\frac{L_1}{L_2}\left(\frac{1}{1 - k^2}\right).$$

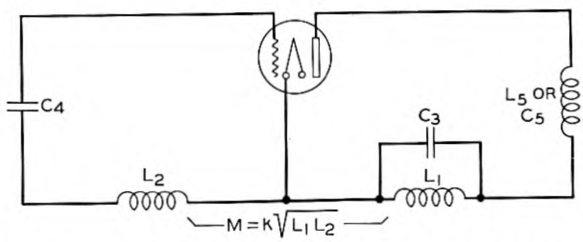


Fig. 13—Reversed feed-back oscillator, plate and grid stabilization.

$$L_5 = L_1\left[1 + \frac{1}{k^2}\left(\frac{L_1C_3}{L_2C_4} - 1\right)\right],$$

$$C_5 = \frac{C_3}{\frac{1}{k^2}\left(1 - \frac{L_1C_3}{L_2C_4}\right) - 1}.$$

impedances to cause the frequency to be independent of changes in battery voltages. In Fig. 11 the stabilizing impedance is placed between the plate and the tuned circuit; in Fig. 12 between the grid and coupling coil; and in Fig. 13 stabilization is accomplished by impedances placed in both positions. Again, the mathematical analysis gives the values of the stabilizing impedances as shown on the figures and indicates that oscillation is possible when these values are used.

Other Types of Oscillator Circuits

As an instance of the stabilizing of another general class of oscillator circuits which has wide application in a number of special cases, attention is drawn to the tuned-plate, tuned-grid circuit of Fig. 14. The

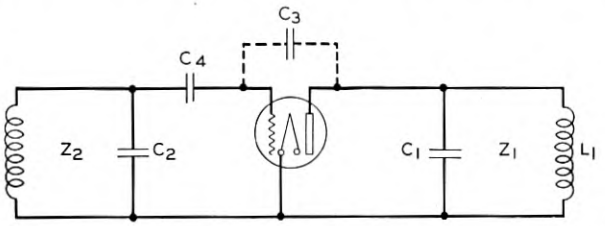


Fig. 14—Tuned-plate, tuned-grid oscillator with no magnetic coupling.

$$C_1 = \frac{L_2}{L_1}\left[C_2 + \frac{(1+\mu)C_3C_4}{C_4 + (1+\mu)C_3} \right] - C_3.$$

input and output circuits are shown in the drawing as consisting of condenser and inductance combinations connected in parallel. At any specified frequency, however, the parallel combination may be replaced

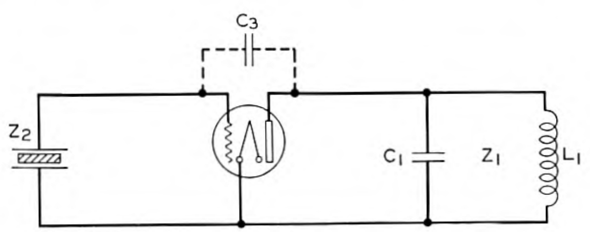


Fig. 14-a.

by a series circuit, or, in fact, by any form of network which has the same impedance, and none of the currents or voltages in the remainder of the circuit will be altered. In particular, the inductance and capacity shown on the input side in Fig. 14 may be replaced by a piezo-electric crystal, as shown in Fig. 14-a, having the same impedance at the operating frequency without affecting the currents and voltages in the remaining parts of the circuit.

It is well known that the frequency of such a piezo-electric oscillator is less affected by changes in battery voltage than is the frequency of the ordinary, nonstabilized electric oscillator. However, the battery voltage does influence the frequency of the piezo-electric oscillator to an extent which is undesirable for certain accurate types of work. It therefore becomes useful to apply stabilization to the piezo-electric oscillator. It will be shown that the stabilization may be accomplished by adjusting the size of the output tuning condenser to such a value that the impedance of the output circuit bears a certain critical relation to the impedance of the crystal, while at the same time, the circuit as a whole fulfills the conditions necessary for the existence of oscillations.

The same kind of stabilization is, of course, applicable to an electric oscillator having analogous relations between the input and output impedances. Thus, it is often possible to stabilize the Hartley oscillator by moving the connection between the filament and coil to different positions on the coil, until that one which gives the proper ratio of input to output impedances has been found. In the case of the Hartley and Colpitts oscillators, however, it is more often preferable to stabilize by the special circuit arrangements illustrated in Figs. 1 to 7, while, on the other hand, the tuned-grid, tuned-plate type of circuit lends itself readily to stabilization by adjustment of the output circuit.

Numerical expressions for the proper impedance relations may be obtained by noting that the circuit of Fig. 14 may be generalized into the circuit of Fig. 1 by regarding Z_4 and Z_5 as zero, while Z_2 comprises the whole input network which may consist of various arrangements of coils, condensers, grid leaks, and the like, and, in a similar fashion, Z_1 comprises the whole output network. The mathematical analysis given in connection with Fig. 1 may therefore be adapted to fit Fig. 14 immediately, and in place of (6) and (7) we have the two expressions:

$$X_0(r_gX_1 + r_pX_2) + \mu r_gX_1X_2 = r_pX_2^2 + r_gX_1^2, \tag{11}$$

$$X_0(r_pr_g - X_1X_2) = - X_1X_2(X_1 + X_2). \tag{12}$$

The requirements of (11) are:

$$r_p = r_g\frac{X_1}{X_2}\left[\frac{\mu X_2 + X_0 - X_1}{X_2 - X_0}\right], \tag{13}$$

which may be used to eliminate r_p in (12) and gives

$$X_0 r_g^2(\mu X_2 + X_0 - X_1) = X_2^2(X_0 - X_1 - X_2)(X_2 - X_0). \quad (14)$$

In order for the frequency to be independent of r_g, it is necessary for one of the factors on the left-hand side of the equation to be zero. This, however, necessitates that one of the factors on the right-hand side of (14) also should be zero. Investigation shows that the only pair of factors of (14) that may both be zero, and still be consistent with (13) is the following:

$$\mu X_2 + X_0 - X_1 = 0, \quad (15)$$

$$X_2 - X_0 = 0. \quad (16)$$

Elimination of X_0 between these two expressions results in the following relation:

$$(1 + \mu)X_2 = X_1. \quad (17)$$

The frequency is then given by the expression:

$$X_1 + X_3 = 0. \quad (18)$$

Equation (17) expresses the relation which is required between the reactances of the input and the output network in order to provide for a constant frequency with varying battery voltages.

In the application of this stabilization to a piezo-electric oscillator such as is shown in Fig. 14-a it sometimes happens that stability improves with decrease in the value of the output tuning capacity but oscillations cease before complete stabilization is secured. The explanation for this and its remedy may be obtained from (17) and (18) by supposing that the reactance, X_2, of the crystal may be represented by an antiresonant circuit, C_2 and L_2, in series with a capacity, C_4, while the output reactance, X_1, consists of the antiresonant circuit, C_1 and L_1. Thus, the value of C_1 which satisfies (17) and (18) is

$$C_1 = \frac{L_2}{L_1}\left[C_2 + \frac{(1 + \mu)C_3}{C_4 + (1 + \mu)C_3} C_4 \right] - C_3. \quad (19)$$

For discussion, the form which (19) takes in the absence of the stopping condenser, C_4, is:

$$C_1 = \frac{L_2}{L_1}[C_2 + (1 + \mu)C_3] - C_3.$$

This shows that a fairly large value of C_1 may be required when C_4 is absent, which places the tuning of the plate antiresonant circuit in a rather critical portion of its reactance characteristic. In order to avoid this, the introduction of a fairly small condenser at C_4 is advantageous. Thus, if C_4 were made somewhat smaller than C_3, then the value of C_1 required by (19) is roughly:

$$C_1 = \frac{L_2}{L_1}[C_2 + C_4] - C_3$$

which gives an appreciably smaller value of C_1 and results in stabilization with a much less critical adjustment than is the case when the stopping condenser is absent.

In all of the above analyses, the requirement of a capacity or an inductance is indicated by the fact that the signs come out right in the final equations. If the wrong type of reactive element were used, it would result, for example, that a negative inductance apparently would be required, which of course would indicate the requirements of a capacitance.

Another Type of Stabilization

A third general type of stabilization may be illustrated by considering a hypothetical oscillator having its plate circuit coupled back to its grid circuit by means of a transformer coil with a coefficient of coupling equal to unity. Methods of obtaining the equivalent effect

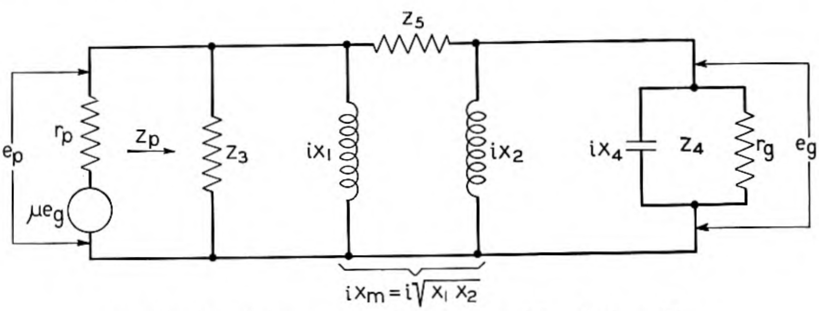


Fig. 15—Equivalent circuit of oscillator with unity coupling.

of such a coil under practical operating conditions will be described later, so that for the present it will be assumed that a unity coupled coil is on hand. The equivalent circuit diagram of the oscillator is shown in Fig. 15, where the primary and secondary windings of the coil are indicated at X_1 and X_2 respectively.

From the properties of unity coupled coils it follows that, no matter what impedances are hung across the coil, or connected between windings, the ratio of the voltage across the secondary to the voltage across the primary depends upon the coil reactances only, and not at all upon the attached impedances. In the circuit of Fig. 15 this ratio is given by the expression:

$$\frac{e_g}{e_p} = -\sqrt{\frac{X_2}{X_1}}. \tag{20}$$

In general, the voltage e_p may be expressed in terms of the impedance looking out of the plate-filament terminals of the tube. Thus

$$e_p = -\frac{\mu e_g Z_p}{r_p + Z_p} \tag{21}$$

where Z_p is the aforementioned impedance.

From (20) and (21) there results:

$$\frac{1}{Z_p} = \frac{1}{r_p}\left[\mu\sqrt{\frac{X_2}{X_1}} - 1\right]. \tag{22}$$

This equation completely expresses the operation of the oscillator in so far as impedance relations for the fundamental current component are concerned. From Fig. 15 ordinary circuit analysis shows that Z_p may be written

$$\frac{1}{Z_p} = \frac{1}{Z_3} + \frac{1}{Z_4}\left(\frac{X_2}{X_1}\right) + \frac{1}{iX_1} + \frac{1}{Z_5}\left(1 + \sqrt{\frac{X_2}{X_1}}\right)^2$$

so that (22) becomes:

$$\frac{1}{Z_3} + \frac{1}{Z_4}\left(\frac{X_2}{X_1}\right) + \frac{1}{iX_1} + \frac{1}{Z_5}\left(1 + \sqrt{\frac{X_2}{X_1}}\right)^2 = \frac{1}{r_p}\left[\mu\sqrt{\frac{X_2}{X_1}} - 1\right]. \tag{23}$$

The next step is to separate this into its real and imaginary components. We stipulate, as in the previous analyses, that the losses in the external circuit elements are small compared with those occasioned by the grid resistance of the tube. With this understanding, Z_4 may be separated into two parts, the one iX_4, in parallel with the other which constitutes the grid resistance r_g. Both Z_3 and Z_5 are taken as pure reactances. Thus the last expression yields the two equations:

$$\frac{1}{X_3} + \frac{1}{X_4}\left(\frac{X_2}{X_1}\right) + \frac{1}{X_1} + \frac{1}{X_5}\left(1 + \sqrt{\frac{X_2}{X_1}}\right)^2 = 0, \tag{24}$$

$$\frac{1}{r_g}\left(\frac{X_2}{X_1}\right) = \frac{1}{r_p}\left[\mu\sqrt{\frac{X_2}{X_1}} - 1\right]. \tag{25}$$

Equation (24) contains neither r_p, r_g, nor μ, so that the important conclusion can be drawn that the frequency of an oscillator with unity coupling between the plate and grid circuits depends only upon the inductances and capacities in the circuit, and not at all upon the tube parameters, r_p, r_g, and μ; provided, however, that the losses in the external circuit are small, and the harmonic voltages across the tube are small enough to allow r_p and r_g to be considered as pure resistances. In this connection, the interelectrode capacities may be grouped with the external circuit elements forming X_3, X_4, and X_5, so that no high-frequency difficulty is to be anticipated from them.

Equation (25) contains the relation between r_p, r_g, and μ necessary to insure the presence of oscillation. In practice, the amplitude of the oscillations builds up until this relation is satisfied.

The foregoing theory of the action of a unity coupled oscillator has led to an extremely useful and desirable result, namely, the independence of frequency and operating voltages. The point now remaining to be shown is how to get the unity coupling.

In attacking this question, the first thing to notice is that our theory does not require that the unity coupling condition,

$$M = \sqrt{L_1 L_2}$$

should be obtained. What actually is required is the much less rigid condition:

$$X_m = \sqrt{X_1 X_2} \tag{26}$$

where X_1 and X_2 are not limited to inductance alone.

Thus, imagine one of the impedances, say X_2, to consist of a coil L_2, in series with a condenser, C_2. We have then, by (26):

$$\omega^2 M^2 = \omega L_1 \left(\omega L_2 - \frac{1}{\omega C_2} \right) \tag{27}$$

or, writing

$$M = k\sqrt{L_1 L_2}$$

where k may now be less than one, we have from (27)

$$C_2 = \frac{1}{\omega^2 L_2 (1 - k^2)} \tag{28}$$

which gives the value of C_2 necessary to provide "unity coupling" at the operating frequency.

The value of X_2/X_1 is thus

$$\frac{X_2}{X_1} = \frac{\omega L_2 - \frac{1}{\omega C_2}}{\omega L_1} = \frac{\omega L_2 - \omega L_2(1 - k^2)}{\omega L_1} = k^2 \frac{L_2}{L_1}. \tag{29}$$

In practice, X_3, X_4, and X_5 would usually be capacities, to correspond to the circuit of Fig. 17. With this arrangement, and the relation given by (29) we have the frequency from (24):

$$\omega^2 = \frac{1}{L_1\left[C_3 + k^2 C_4 \frac{L_2}{L_1} + C_5 \left(1 + k\sqrt{\frac{L_2}{L_1}} \right)^2 \right]} \cdot \qquad (30)$$

The value of C_2 is thus written from (28) and (29) as follows:

$$C_2 = \left(\frac{L_1}{L_2}\right)\left[\frac{C_3 + k^2 C_4 \frac{L_2}{L_1} + C_5 \left(1 + k\sqrt{\frac{L_2}{L_1}} \right)^2}{1 - k^2}\right]. \qquad (31)$$

This is the general value of C_2 needed to stabilize the oscillator, and applies to any grid-stabilized oscillator where the unity coupling concept can be employed. In the case where C_5 and C_4 are small enough to be neglected we have the equivalent circuit of the reversed feed-back oscillator of Fig. 12, and for the value of the stabilizing capacity:

$$C_2 = \frac{L_1}{L_2}\frac{C_3}{1 - k^2} \cdot \qquad (32)$$

When the notation of Fig. 17 is reconciled with that of Fig. 12, this is in agreement with the conclusion reached for the reversed feed-back oscillator by the former method of analysis.

The present analysis has the twofold advantage of allowing the interelectrode capacities to be included, which results in (31) instead of (32); and of giving a more readily interpreted picture of the relation required for stability, namely the "unity coupling" condition of equation (26). Equation (31) is moreover applicable to the tuned-plate, tuned-grid type of oscillator, when there is magnetic coupling between the input and output circuits. Thus, in the particular instance when L_1 and L_2 are equal, as also are C_3 and C_4, we have from (31):

$$C_2 = C_3 \frac{(1 + k^2)}{(1 - k^2)} + C_5 \frac{(1 + k)}{(1 - k)} \cdot \qquad (33)$$

Hence, if tuning is done by "ganging" C_3 and C_4 together and varying them simultaneously, the stability may be maintained for all frequencies by making C_2 to consist of two parts: the one a fixed capacity equal to

$$C_5 \frac{(1 + k)}{(1 - k)}$$

and the other a variable capacity "ganged" together with C_3 and equal to

$$C_3 \frac{(1+k^2)}{(1-k^2)}\cdot$$

In order to insure fulfillment of the requirements of this theory, it was mentioned above that the oscillator should be relatively free from harmonics, so that r_p and r_g may be taken as pure resistances. That this requirement can be successfully met may be demonstrated by reference to Fig. 15 and consideration of the means which would be employed if the circuit represented an amplifier with e_g impressed on the grid of a following tube instead of on the grid of the driving tube itself. In such an arrangement, it is well known that the distortion is least when the impedance looking out of the plate of the driving tube is made materially larger than the internal plate resistance of the tube itself; and, second, when the impedance looking back out of the grid of the driven tube is made materially smaller than the internal grid resistance of the tube itself. The conditions which determine how nearly these two requirements may be met in the oscillator tube are governed by (22). Thus

$$\frac{r_p}{Z_p} = \mu\sqrt{\frac{X_2}{X_1}} - 1. \tag{34}$$

This should be small in order that the first requirement mentioned above should be fulfilled. Hence $\mu k\sqrt{L_2/L_1}$ should exceed unity by as little as is consistent with reliable oscillation. When we come to consider the second of the requirements for decreasing harmonics we obtain an expression analogous to (34), namely:

$$\frac{Z_g}{r_g} = \mu\sqrt{\frac{X_2}{X_1}} - 1 \tag{35}$$

so that the requirement that the impedance looking back out of the grid should be less than the grid resistance is satisfied by the same condition as that required by (34); namely, that $\mu k\sqrt{L_2/L_1}$ should exceed unity by as little as is consistent with reliable oscillation.

From Fig. 17, by regarding C_3 and C_5 as zero, the feed-back type of oscillator is obtained. Again, when C_3 and C_4 are zero, the Hartley oscillator results. The stabilization both of the feed-back oscillator and the Hartley type by the method shown on the figure was not described in the analysis of Figs. 3 and 9 inasmuch as the present method places the stabilizing element directly in the tuned circuit, whereas the former method placed it between the tuned circuit and the vacuum tube.

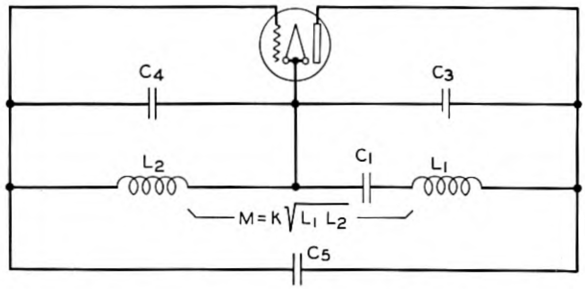


Fig. 16—General circuit of oscillator with unity coupling, plate stabilization.

$$C_1 = \frac{L_2}{L_1}\left[\frac{C_3k^2\frac{L_1}{L_2} + C_4 + C_5\left(1 + k\sqrt{\frac{L_1}{L_2}}\right)^2}{1 - k^2}\right].$$

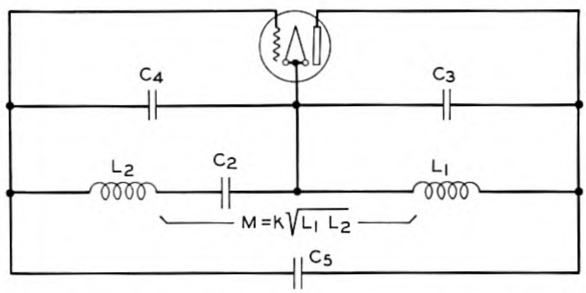


Fig. 17—General circuit of oscillator with unity coupling, grid stabilization.

$$C_2 = \left(\frac{L_1}{L_2}\right)\left[\frac{C_3 + k^2C_4\left(\frac{L_2}{L_1}\right) + C_5\left(1 + k\sqrt{\frac{L_2}{L_1}}\right)^2}{1 - k^2}\right].$$

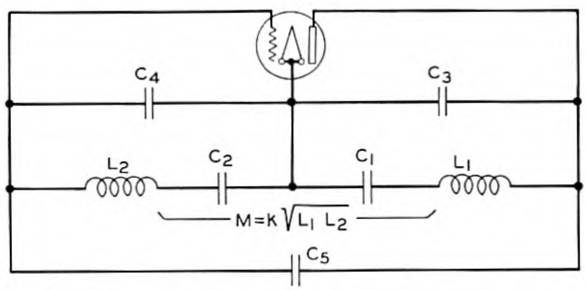


Fig. 18—General circuit of oscillator with unity coupling, plate and grid stabilization.

$$\omega^2k^2L_1L_2 = \left(\omega L_1 - \frac{1}{\omega C_1}\right)\left(\omega L_2 - \frac{1}{\omega C_2}\right).$$

Besides the general circuit of Fig. 17, where stabilization is accomplished by imposing a critical value on C_2, there is the alternative shown in Fig. 16, where the stabilizing element is C_1, a condenser located in series with L_1, and of such a size as to cancel the leakage reactance between L_1 and L_2. The formula giving this size in terms of the other circuit constants is shown in the figure, and the condition that the harmonic content should be small for circuits of the general type of Fig. 16 is that $(\mu/k)\sqrt{L_2/L_1}$ should exceed unity by as little as is consistent with reliable oscillation.

From Fig. 16 the circuits of various types of oscillators may be derived in the manner which was described in connection with Fig. 17. Of these circuits, the reversed feed-back and the Hartley were not described in connection with Figs. 11 and 12.

A combination of the features of Figs. 16 and 17 may be employed in the manner shown in Fig. 18 where condensers C_1 and C_2 are placed in series both with L_1 and L_2, respectively. In this case, the formula for the required size of C_1 and C_2 becomes quite cumbersome when expressed in terms of the other circuit elements, only. Since, however, the frequency is usually known approximately, we may use the relation

$$X_m^2 = X_1 X_2$$

for finding C_1 and C_2 in terms of ω and get:

$$\omega^2 k^2 L_1 L_2 = \left(\omega L_1 - \frac{1}{\omega C_1}\right)\left(\omega L_2 - \frac{1}{\omega C_2}\right). \tag{36}$$

As in the case of Figs. 16 and 17, so also may the circuit of Fig. 18 be modified to correspond to the reversed feed-back, the feed-back and the Hartley types of oscillators where stabilization is accomplished both on the plate and on the grid sides.

Experiment

Of course, Figs. 1 to 18 are intended to represent only the fundamentals of the corresponding circuits. For practical operation these circuits would have to include the usual stopping condensers, leak resistances, sources, choke coils, and accessories. These circuit elements should be so valued and introduced into the oscillator circuit as a whole as not to interfere with the relations required by the analyses, in order to maintain the stabilizing effects of the stabilizing impedances. As to the choke coils, this means merely that they must be what the name implies, that is, a substantially infinite impedance. In the case of a Hartley-type oscillator, where the reactance is chosen to be located in the grid leads instead of in the plate leads, a condenser must be used. This may replace the conventional stopping condenser. Where the reactance is in the plate lead for a similar type of oscillator, the stopping condenser in the grid lead should be large so as to have negligible impedance. Similar expedients are suggested for the impedances of the other types of oscillator circuits.

As typical of the general method whereby any of the simplified circuits of Figs. 1 to 18 may be elaborated into a conventional circuit of this kind, including the various adjunctive circuits, Fig. 19 should be

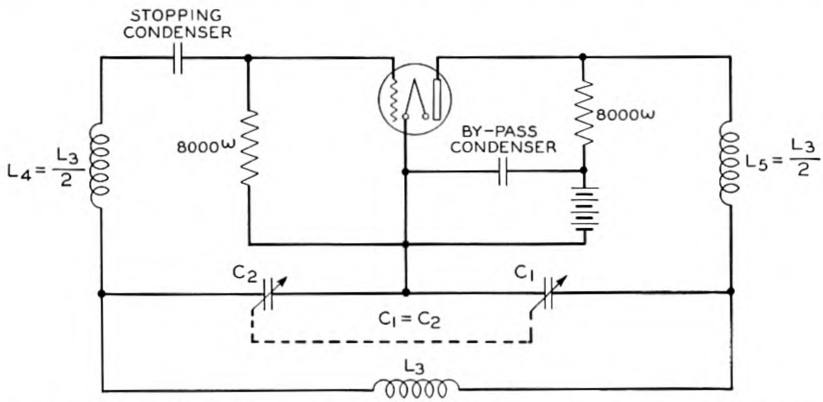


Fig. 19—Practical arrangement of oscillator with three series tuned coupled circuits.

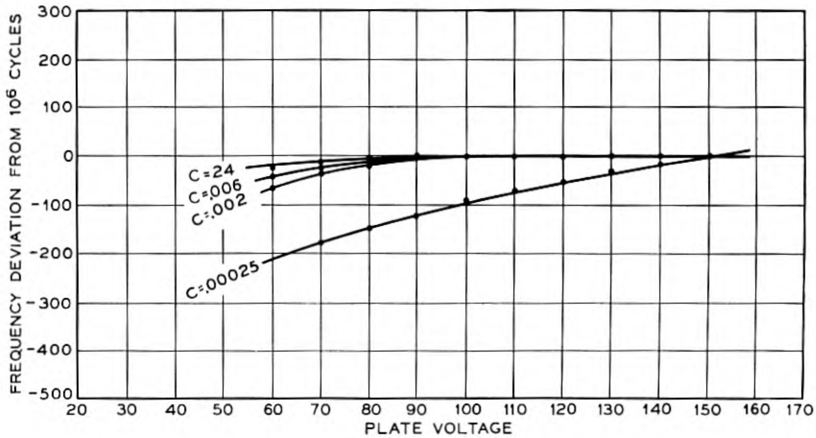
referred to. This figure illustrates a complete wiring diagram of the oscillator of Fig. 7 and shows an example of stabilization by means of the inductance L_5 in the plate circuit and the inductance L_4 in the grid circuit. In addition to satisfying the relation shown in Fig. 7, it may be noticed that the value of L_5 is such as to tune with C_1 to the oscillation frequency, and, similarly, the value of L_4 is such as to tune with C_2 to the oscillation frequency. Under such conditions a resistance of appreciable value may be introduced into the circuit of L_3 without affecting the frequency of the stabilization. The reason for this may be explained briefly as follows:

Consider a single series circuit formed of one of the three meshes of Fig. 19, for instance that composed of the elements, r_g, in parallel with the 8000-ohm leak, L_4, and C_2. This circuit is in series resonance at the frequency at which the circuit as a whole oscillates. Therefore it tends to introduce resistance impedance only into whatever circuits it is reactively coupled with. Thus, the effect of this circuit upon the adjacent circuit, L_3, C_1, C_2, with which it is coupled is to introduce resistance only. Similarly, if this last circuit operates at series resonance, only resistance is introduced into the plate circuit, L_5, C_1, and r_p, in parallel with the d-c feed of 8000 ohms, with which it is coupled. Hence, if the plate circuit likewise operates at series resonance, a change in resistance of any part of the circuit will change only the resistance into which the tube works and therefore will leave the frequency unaltered.

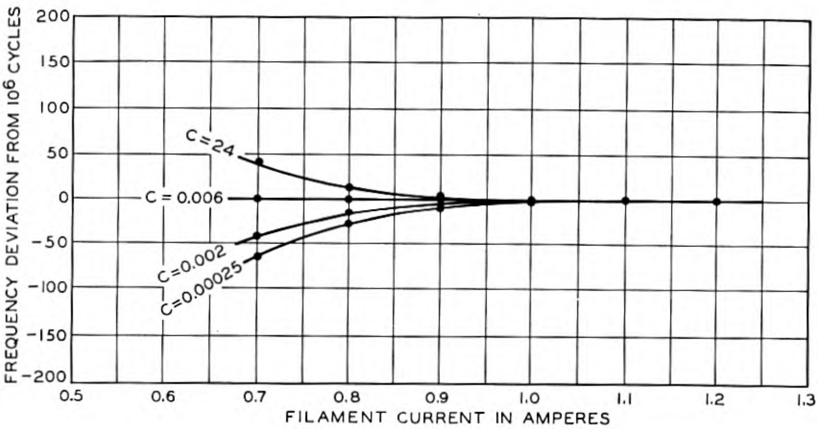
In a more general sense, any of the oscillator forms discussed may be stabilized even when the resistance in the external circuit is not inappreciable, the effect of the external resistance manifesting itself in two ways: first, a value of stabilizing reactance slightly different from that given in the above formulas may be required, and second, the frequency, instead of being absolutely independent of battery voltage variations, goes through a maximum or a minimum as the battery voltages change, the voltage at which this maximum or minimum occurs depending upon the exact value of the stabilizing reactance. An exact mathematical analysis of this more general case yields formulas for the stabilizing reactances which involve r_g or r_p and hence are not as useful even in a case where the resistance in the external circuit is of importance as are the formulas presented above. The latter may be used as first approximations in any event.

In practice it has been found that when ordinary precautions are taken to insure a low-loss external circuit, the relations given above hold very accurately and any variations in frequency then existing as a result of varying battery voltages may be traced to either one of two causes, both of which may be guarded against: first, the interelectrode capacities of the tube may be sufficient to enter into the impedance relations. In this event, a change in the form of circuit, such as the use of the tuned plate-tuned grid arrangement of Fig. 17, where the interelectrode capacities form a part of the external circuit, will eliminate the difficulty. Second, the harmonic currents caused by the nonlinear characteristics of the vacuum tube may introduce the effect of a reactive impedance back into the fundamental which may vary with battery voltage and so change the frequency. The remedy for this is to provide a low reactance path for the harmonics so that they have no opportunity to build up a reactive voltage across the tube, and also to use grid leaks and other such well-known devices for reducing the harmonic currents generated by the tube.

For the purpose of providing information as to the order of stability which may be expected from the several methods of stabilization outlined above, various quantitative experimental tests have been conducted. The general results of these tests may be summarized by saying that a close adherence to the theoretical requirements results in an oscillator whose frequency depends upon operating voltages to such a small extent that temperature effects become the predominating influence and special precautions must be taken in order to eliminate



(a)



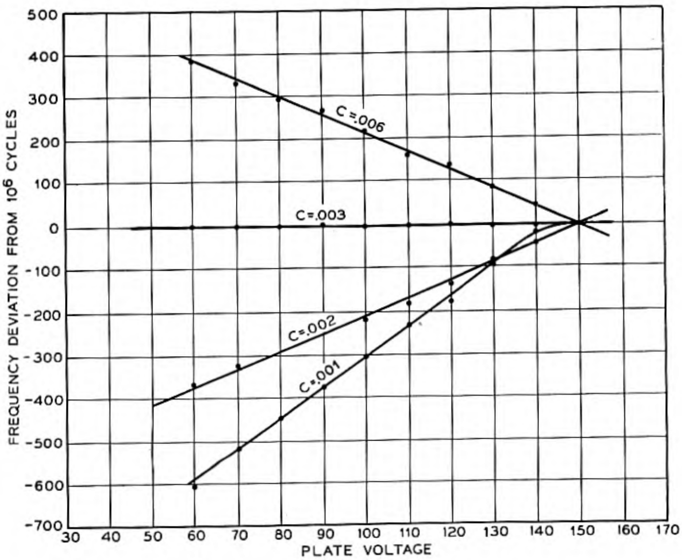
(b)

Fig. 20—Performance curves of reversed feed-back oscillator with tight coupling, grid stabilization.

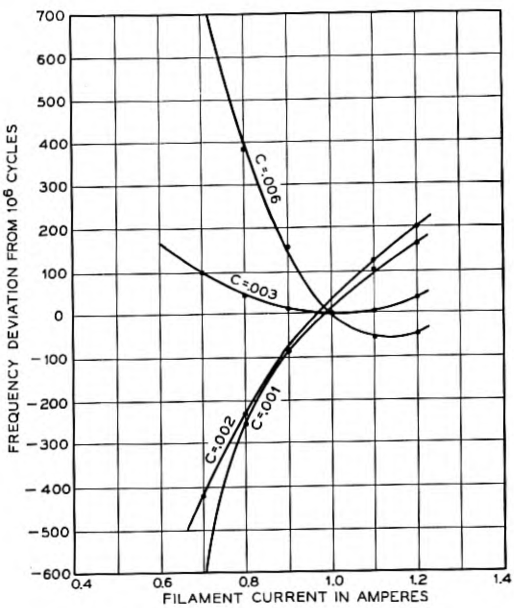
(a) Variation of frequency with plate potential.
(b) Variation of frequency with filament current.

them before data showing the dependence of frequency on operating voltages can be obtained.

For instance, the data for the curves shown in the accompanying Figs. 20 and 21 were secured by the following procedure: A plate



(a)



(b)

Fig. 21—Performance of reversed feed-back oscillator with loose coupling, grid stabilization.

(a) Variation of frequency with plate potential.
(b) Variation of frequency with filament current.

potential of 150 volts and a filament current of one ampere were selected as reference points. The frequency under these conditions was noted. A change to a different operating condition, say 140 volts plate potential and 1 ampere filament current, was made and the frequency measured as rapidly as possible, whereupon the operating voltages were returned immediately to their reference values and the frequency rechecked. Special care was taken to keep the room temperature very constant, but, even so, the heating of the parts of the oscillator circuit by the operating alternating and direct currents was sufficient to affect the frequency to an undesirable extent, requiring that the readings be taken with unusual rapidity in order to return the voltages to normal before the changed operating currents could appreciably affect the temperature of the coils, tube elements, and other parts of the circuit.

The final results, however, are consistent enough to be representative of what can be accomplished, and the two sets of curves shown in the figures bring out a result which was found to hold throughout the investigation; namely that the higher the coefficient of coupling in the coil used to secure feed-back, the less critical was the value of the stabilizing impedance. Thus, in Fig. 20 the coupling was as tight as it was possible to produce by winding the primary and secondary simultaneously upon a tube to form a single layer solenoid, while the coil used for Fig. 21 was made by removing about half the turns from the secondary of the same coil, thus providing for a step-down in voltage as well as a decrease in coupling. A possible explanation for the less critical adjustment required with tight coupling lies in the fact that the tightly coupled coil satisfies the condition for "unity coupling" as given by (26) over a range of frequencies whereas the loosely coupled coil satisfies the condition at only the frequency critically determined by the stabilizing capacity.

This would appear to indicate that the type of stabilization described in the theoretical part of the paper under the unity coupling concept offers certain practical advantages over those types where the stabilizing element is placed between the tuned circuit and the tube, as in the Colpitts oscillator of Figs. 5, 6, and 7, for example, where no magnetic coupling whatever is employed. Experiments with the Colpitts circuit have shown, however, that when the capacities in the tuned circuit are made large relative to the inductance, a very satisfactory degree of stabilization may be secured at the lower frequencies where interelectrode capacities may be neglected. A close inspection of the theory of stabilizing the frequency of the Colpitts oscillator shows an argument analogous to that of the coupled coils, namely that the smaller we make the inductance in the tuned circuit, and the larger we make the capacity, the frequency being kept constant, the greater will be the range of frequencies over which the condition, required by (8), that the series reactance of the tuned circuit be zero, is satisfied to an approximation sufficiently good for practical purposes, and hence the less critical will be the adjustment of the stabilizing impedance.

For the data in Figs. 20 and 21 a reversed feed-back type of oscillator was employed, having an elementary circuit similar to that of Fig. 12. A grid leak was placed in parallel with the stabilizing capacity, but care was taken to see that the resistance of the leak was always so high that its value did not affect the frequency. This was done by using a variable resistance and increasing its value until the frequency no longer changed. A large grid leak has always been found advantageous in securing constancy of frequency, but the size of the leak reaches a practical limitation determined by the time constant which produces the familiar "blocking."

The plate battery potential was fed through a choke in series with the plate inductance coil, the combination of choke and B battery being thoroughly by-passed with a large condenser.

For the purpose of checking the size of the stabilizing capacity to find its agreement with theory, an indirect method was used. The theory requires that

$$C_4 = C_3 \frac{L_1}{L_2}\left(\frac{1}{1-k^2}\right)$$

as shown in Fig. 12. The operating frequency was one megacycle, which made a direct measurement of the coupling coefficient, k, somewhat awkward, so that a method based on the "unity coupling" concept was employed. Thus from (26)

$$k^2\omega^2 L_1 L_2 = \omega L_1\left(\omega L_2 - \frac{1}{\omega C_4}\right) \tag{37}$$

may also be used for determining C_4. The primary, L_1, of the coil was connected through an impedance to a source of e.m.f. of one megacycle, and a vacuum tube voltmeter was placed across it. The condenser, C_4, was placed directly in parallel with the secondary, L_2. With this arrangement the impedance looking into the primary is

$$Z = i\omega L_1 + \frac{\omega^2 k^2 L_1 L_2}{i\left(\omega L_2 - \frac{1}{\omega C_4}\right)}$$

unless points very near the resonance point are considered. This last equation may be written

$$iZ\left(\omega L_2 - \frac{1}{\omega C_4}\right) = -\,\omega L_1\left(\omega L_2 - \frac{1}{\omega C_4}\right) + \omega^2 k^2 L_1 L_2. \qquad (38)$$

The condenser, C_4, was varied until the reading of the vacuum tube voltmeter became zero. This means that Z was zero, and hence (38) gives the value of C_4 required by (37) which is in turn the value needed to stabilize the oscillator.

How well this checked the actual values needed for the case shown in Figs. 20 and 21 may be seen by the following: For Fig. 21 the value of C_4 measured as above, was 4000 $\mu\mu$f. The experimental value was 3000 $\mu\mu$f. For Fig. 20 the measurement gave a rather broad zero on the vacuum tube voltmeter, which was, however, estimated at 8400 $\mu\mu$f. For a check, a measurement was made at 7 megacycles which gave a sharper zero, and a value of 120 $\mu\mu$f. This must be reduced to its equivalent value of 1 megacycle by multiplying by 7^2 which gives 5880 $\mu\mu$f. The experimental curves of Fig. 20 show a noncritical value of 6000 $\mu\mu$f. which is nevertheless in good accord with the above measurements, while in Fig. 21 the agreement is somewhat more striking.

As an example of stabilization of oscillators in the altogether different frequency region from 7 to 40 kc, the following table was taken from data kindly supplied by F. J. Rassmussen:

TABLE I

Frequency kc	L_1 mh	L_2 mh	Coupling k	Stabilizing Capacity μf	
				Experimental	Theoretical
12.5	3.176	6.891	0.131	0.0021	0.0022
40.0	3.210	7.010	0.134	0.0018	0.0023
12.5	3.178	6.974	0.129	0.0021	0.0022
40.0	3.211	7.094	0.133	0.0018	0.0023
7	2.498	0.500	0.720	0.4 to ∞	2.07
15	2.499	0.498	0.723	0.2 to ∞	0.46
15	1.409	0.161	0.695	0.2 to ∞	1.33
23	1.407	0.161	0.698	0.1 to ∞	0.58
23	0.665	0.093	0.677	0.1 to ∞	0.99
31	0.665	0.093	0.681	0.1 to ∞	0.57
31	0.487	0.076	0.663	0.1 to ∞	0.64

This table again emphasizes the less critical adjustment required when the coupling is tight, as in the last seven rows.

It is hoped that the foregoing data and comments will serve as a guide to design methods for constant frequency oscillators, in so far as dependence of the frequency on operating voltages is concerned. Combinations and permutations of the various circuits dealt with will occur to the designer who requires special arrangements to fit special cases. The generalized circuit of Fig. 18 is suggested as being adaptable to meet the most widely varying conditions. This is particularly true at very high frequencies, since all the interelectrode capacities are included in the circuit of that figure.

The popular "push-pull" type of circuit may likewise be generalized to correspond to several of the fundamental circuits illustrated in the figures, and may be stabilized by the methods indicated. However, because of the nonuniformity of vacuum tubes and the added complication of the circuit, no advantage has been obtained by its use, so that the single tube circuits are to be preferred wherever special conditions do not require the push-pull type.

Appendix

The complete and rigorous mathematical relations for oscillation circuits containing vacuum tubes have seldom been discussed in connection with their practical application to useful circuits. In the case of the stabilization of oscillators against changes in battery voltages it is important to base the theory upon as strictly rigorous a mathematical foundation as possible, yet at the same time to be able to express the results in readily useful terms. It will be shown that this desirable result may be attained by a proper interpretation of the meaning of the internal impedances, r_p and r_g, of the vacuum tube.

To show this in the shortest and most obvious way, a simple series circuit will be considered. Let the circuit consist of a resistance, R, a condenser, C, and an inductance, L, all connected in series with a vacuum tube which may be taken as having a "negative resistance" characteristic. In order to increase the generality of the demonstration, a sinusoidal driving voltage, E, of angular frequency, ω, is also allowed to act on the circuit. By Kirchkoff's Law, the current in the circuit is expressed by the equation

$$E = RI + L\frac{dI}{dt} + \frac{1}{C}\int Idt + V \tag{1}$$

where V is the drop across the vacuum tube. As a general expression for V in terms of the current the following expression may be used:

$$V = V_0 + A_1I + A_2I^2 + A_3I^3 + \cdots. \tag{2}$$

We are interested in the "steady state" solution, and accordingly a Fourier series will be the most general form which can be assumed for the current. It is convenient to write the series in the following form:

$$I = \sum_{n=-\infty}^{\infty} \frac{b_n}{2} \epsilon^{\iota n\omega t} \tag{3}$$

or, for brevity, in the symbolic form

$$I = \sum I(n\omega) \tag{4}$$

where the summation is understood to extend from minus infinity to plus infinity. Substitution of (4) and (2) into (1) gives:

$$\begin{aligned} E = R\sum I(n\omega) + L\sum in\omega I(n\omega) + \frac{1}{C}\sum \frac{I(n\omega)}{in\omega} \\ + V_0 + A_1\sum I(n\omega) + A_2\sum\sum I(n\omega)I(m\omega) \\ + A_3\sum\sum\sum I(n\omega)I(m\omega)I(l\omega) + \cdots. \end{aligned} \tag{5}$$

For the component of fundamental frequency, ω, we get

$$\begin{aligned} E(\omega) = RI(\omega) + Li\omega I(\omega) + \frac{I(\omega)}{Ci\omega} + A_1 I(\omega) \\ + A_2 \sum_{n+m=1} I(n\omega)I(m\omega) + A_3 \sum_{n+m+l=1} I(n\omega)I(m\omega)I(l\omega) + \cdots \end{aligned} \tag{6}$$

where the summation terms involve the products of all frequency components which beat together to give the fundamental, as indicated. In order to put the last expression in symmetrical form, it is convenient to multiply and divide each of the summation terms by $I(\omega)$ so that we may write

$$\begin{aligned} E(\omega) = I(\omega)\Big[R + Li\omega + \frac{1}{Ci\omega} + A_1 + \frac{A_2}{I(\omega)} \sum I(n\omega)I(m\omega) \\ + \frac{A_3}{I(\omega)} \sum I(n\omega)I(m\omega)I(l\omega) + \cdots \Big]. \end{aligned} \tag{7}$$

This expression exhibits the terms in square brackets in the form of an impedance, and shows that the vacuum tube may be treated as an ordinary linear circuit element if it is considered as having the impedance

$$Z = A_1 + \frac{A_2}{I(\omega)} \sum I(n\omega)I(m\omega) + \frac{A_3}{I(\omega)} \sum I(n\omega)I(m\omega)I(l\omega) + \cdots. \tag{8}$$

Of course, the numerical value of such an impedance cannot be found from this expression alone, but in oscillator analysis there is no necessity for its numerical evaluation. The very important fact that the nonlinear elements in a circuit network may be replaced by equivalent impedances so that the ordinary circuit analysis can be employed has been demonstrated. It is possible to tell something about the form of the tube impedance from (8). Thus, the first term, namely A_1, is a real quantity and contributes a part of the total effective resistance of the tube. All of the remaining terms are, in general, complex, depending upon the phases of the different harmonic currents. Thus, the conclusion is reached that a nonlinear resistance may be reduced to an equivalent linear impedance, but that this impedance has a reactive as well as a resistive component in the general case. There is at least one important instance where the equivalent linear impedance is resistive only. This occurs when the impedance in the circuit external to the vacuum tube contains resistance, only, to all of the harmonic currents.

With the general conception of the impedance of the vacuum tube, described above, the fundamental component of (1) becomes

$$E = \left[R + i\omega L + \frac{i}{i\omega C} + r + iX \right] I \tag{9}$$

where the tube impedance is represented by $r + iX$.

When the driving voltage, E, is zero, as in the case of oscillators, then for a finite current to exist the oscillation conditions are:

$$\left.\begin{aligned} R + r &= 0 \\ \omega L - \frac{1}{\omega C} + X &= 0 \end{aligned}\right| . \tag{10}$$

In the treatment of oscillator networks employed in the foregoing paper the quantities, r_p and r_g, are used in the sense of the resistance, r, in (9) and (10) of this appendix. The reactive component, X, of the tube impedance has been neglected in the paper, for the reason that all of the circuits discussed are of such character that the reactance of the external circuit to the harmonic currents may be made quite low, and the nonlinearity of the vacuum tube characteristics is not such as to cause excessive production of harmonics.

In the case of the dynatron type of oscillator, where the harmonic currents are especially strong, it has been found by experiment that the reactive component of the tube impedance cannot be neglected, but that it is, in fact, altogether responsible for the variation in frequency with battery voltages which is characteristic of the dynatron oscillator.

Some Physical Properties of Wiping Solders *

By D. A. McLEAN, R. L. PEEK, Jr., and E. E. SCHUMACHER

The plasticity of a number of solders at wiping temperatures has been determined by compression tests between parallel plates. The character of the flow is found to be that corresponding to a linear relation between shearing stress and a fractional power of the velocity gradient. This corresponds approximately to a relation between rate of compression (dh/dt) and sample height (h) given by the equation: $dh/dt = kh^b$, in which k and b are constants, of which b is independent of the test conditions. For viscous materials $b = 5.0$; for most solders b is greater than 5.0, and increasing values of b are associated with lower temperature gradients of plasticity. It is shown that a solder must have a low temperature gradient of plasticity in order to be properly wiped, and that determination of the value of b by means of a plasticity test can therefore be used to evaluate the working properties of a solder.

A number of factors upon which the plasticity of wiping solders and the porosity of wiped joints may depend have been investigated. In particular, it is shown that segregation is not responsible for porosity, but that the latter may be dependent upon the particle size of the solid phase at wiping temperatures. The relation of particle size to the wetting power of the liquid phase is discussed.

A WIPING solder is one used in joining sections of lead (or lead alloy) piping or sheathing, such as that used for telephone cables. As such, it must wet the sheathing readily, must be coherent and plastic enough to be worked with the hands over a considerable temperature range, and must form a strong, non-porous joint. The desirable qualities for a wiping solder have been more fully enumerated by Schumacher and Basch,[1] but the above will suffice for consideration in connection with this paper.

One phase of a general investigation of wiping solders has been a study of their plastic properties in the temperature range in which they are wiped. In this study consideration has been given not only to the working properties which a solder must have in order to be properly wiped, but also to certain other properties which may affect the character of the wiped joint, particularly those which may result in porosity of the joint.

The temperatures at which the tests described below were made lie in the wiping range, and are intermediate to those represented by the solidus and liquidus lines in the respective equilibrium diagrams of the solders investigated. At such temperatures the alloys form two phase

* Presented before the Society of Rheology, Rochester, New York, December 28, 1931. Published in *Jour. of Rheology*, January, 1931.

[1] E. E. Schumacher and E. J. Basch, *Ind. & Eng. Chem.*, **21**, 16 (1929).

systems, in which a solid phase, composed principally of one constituent, is dispersed in a liquid phase, whose composition and abundance is a function of the temperature and the gross composition of the alloy. These facts suggest that the character of a solder may be related in the wiping range to the following properties: solid-liquid ratio, solid particle size, viscosity of the liquid phase, tendency to segregation, and the interfacial tension between solid and liquid phases.

Objects of Investigation

The objects of the investigation may be summarized under the following headings:

1. Character of Flow: It was hoped that by means of plasticity determinations, data would be obtained permitting the flow of the various alloys to be formulated in simple terms, and that appropriate flow constants could be evaluated which would suffice to describe the nature and extent of deformation under specified conditions of stress. This object was largely realized through the discovery that the flow is similar in character to that of many colloidal dispersions, corresponding approximately to what would be expected if a power relation exists between stress and velocity gradient.

2. Relation of Plasticity to Workability: The primary purpose of the plasticity studies was to determine if the information thus obtained could be related to the working characteristics of the solders. If a clear-cut relation could be found, the plastometer could be employed in testing new alloys for use as solders, and in making control tests on solders as supplied to the splicers. As a matter of fact, a simple relation between the plasticity data and workability was found to exist, on which can be based a method of employing the plastometer in research studies of solders. In addition the plasticity appears to be very sensitive to composition changes, and its measurement with the plastometer should in consequence be useful as a control test. Whether this latter possibility can be realized in engineering practice has not, as yet, been ascertained.

3. Plasticity and Solid-Liquid Ratio: Early in the investigation it was suggested that the plasticity might be related to the solid-liquid ratio. If this were so, the plasticity and its temperature gradient could be predicted from the equilibrium diagram for any particular alloy, and calculation of solid-liquid ratios could be used to supplant or to corroborate plasticity determinations. While for any one solder, of course, the plasticity increases with the ratio of liquid to solid, no relation common to any group of solders was found to hold between plasticity and the solid-liquid ratio.

4. Porosity: At the start of the investigation it was thought that the porosity of the joint might be related to the plasticity of the solder. As this was not found to be the case, other factors which might be responsible for this condition were investigated. In particular, experiments were performed to determine if porosity is dependent upon segregation, or if it is related to the particle size of the solid phase at wiping temperatures. Porosity and segregation were found to be quite independent, but a relation between porosity and particle size was found in the two tests made in this connection.

Experimental work directed toward these objects has involved the following groups of experiments:

1. Plasticity studies.
2. Segregation studies.
3. Investigation of particle size.

The first of these groups was planned to cover the first three objects enumerated above, while the other two groups are of interest in connection with porosity.

Alloys Investigated

Six different alloys were tested as described in Table I. In this table, the compositions sought in preparing the samples are given, together with the composition found by analysis at the top and bottom of the cast (T & B), except for solders Nos. 5 and 6, of which no analyses were made.

Nos. 2 and 3 are good, workable solders giving non-porous joints, with No. 3 possibly being given preference in ease of handling while the joint is being wiped. No. 1 is also satisfactory from the standpoint of workability, but the joints formed from it are often porous, while the splicers described it as being somewhat " coarse." No. 4 is unsatisfactory in all respects. Some tests on No. 5 showed it to be good, but in recent tests the workability has been poor and a number of porous joints have been found. Solder No. 6 seemed to be fair both in workability and porosity, but the number of tests has been altogether too limited for definite conclusions to be drawn. The workability of neither of these last two solders is as good as that of the lead-tin and lead-tin-cadmium alloys. The results of wiping and porosity tests are summarized to the right of Table I.

A point which the authors wish to stress is that they are interested here only in two of a number of factors which may affect the utility of solders; namely, workability and porosity. Therefore, that a solder is good in those respects does not necessarily imply that it is a good solder in a general sense, and leaves open the question as to

TABLE I

	Per Cent Pb	Per Cent Sn	Per Cent Cd	Per Cent Bi	Per Cent Zn	Workability	Lack of Porosity	Tested at ° C.
Solder No. 1						Good	Porous	196, 204 and 210
Desired	65	35						
Actual T		34.50						
Actual B		34.62						
Solder No. 2						Good	Good	183, 185 and 194
Desired	60	40						
Actual T		40.32						
Actual B		40.69						
Solder No. 3						Excellent	Good	170, 184 and 192
Desired	67	24	9					
Actual T		24.00	9.28					
Actual B		24.13	9.38					
Solder No. 4						Poor	Porous	
Desired	60			40				
Actual T				40.25				
Actual B				40.20				
Solder No. 5						Fair	Porous	149,153, 157, 161
Desired	67		4.25	28.5	.25			
Solder No. 6						Fair	Good	140, 148, 156, 160
Desired	65	13		22				

T = Top of cast.
B = Bottom of cast.

whether or not consideration of other factors may show it to be unsatisfactory as a wiping solder.

Plasticity Studies

Experimental Procedure

The plastometer used in the investigation is a modification of the instrument used by Williams [2] in studies on rubber compounds. It has been fully described elsewhere [3] by one of the writers. A heat-insulated cylindrical steel block provided with heating elements and thermocouples contains a central cylindrical well 2½ inches (6.35 cm.) in diameter and 5 inches (12.70 cm.) deep. The block itself is about 12 inches (30.48 cm.) in diameter, being made large to prevent rapid temperature fluctuations. A flat-bottomed cylindrical plunger, also carrying heating elements and a thermocouple, fits into the well with a small clearance. The load on the plunger may be adjusted to suit the conditions of test. A sample of the material to be tested is placed under the plunger, given 90 minutes to come to temperature and the plunger released and allowed to compress the sample. The

[2] I. Williams, *Ind. & Eng. Chem.*, **16**, 262 (1924).
[3] R. L. Peek, Jr., "Parallel Plate Plastometry" (now being prepared for publication).

temperature is frequently checked, and can be controlled within $\pm$ 1° C. by adjusting external resistances. The sample height is measured at intervals by an Ames gauge which moves with the plunger, a complete record of sample height vs. time being thus obtained. With one exception, the tests here reported were continued for 30 minutes.

The sample size used was one which has been found convenient—a cylinder 1.374 inches (3.490 cm.) in diameter and .300 inch (.762 cm.) high. Samples were obtained from a single cast of each solder, tested for blow-holes and inclusions by density measurements. Samples were milled to dimensions with a tolerance of $\pm$ .001 inch (.0025 cm.). The weight of the plunger and load was 30 lbs. (13.63 kg.).

Character of the Flow

The data directly obtained in a typical run, readings of sample height vs. time from start of run, are plotted in Fig. 1, together with tangents to the curve fitting the points plotted. By computing the slope of such tangents there are obtained values of the rate of compression, dh/dt, corresponding to various values of sample height, h. If these are plotted logarithmically, log dh/dt vs. log h, as in Fig. 2, a straight line is obtained. This shows that under the test conditions employed, the relation between sample height and rate of compression is of the form:

$$\frac{dh}{dt} = kh^b, \tag{1}$$

where k and b are empirical constants. All the tests made on solders gave results that could be fitted with an equation of this type. At any one temperature the value of b was constant for a given solder, while the value of k varied with the load and sample volume but was independent of the initial sample height. A limited number of tests have indicated that all runs made on any one solder at a given temperature can be represented by an equation of the type:

$$\frac{dh}{dt} = K\frac{W^a h^b}{V^c}, \tag{2}$$

where W is the load, V the sample volume and K, a, b, and c are constants characteristic of the material.

It has been shown both theoretically [4] and experimentally [3, 5] that

[4] O. Reynolds, *Phil. Trans., Lond.*, **177A**, 157 (1886).
[5] Ormandy, "The Engineer," **143**, 362, 393 (1927).

a cylindrical sample of a viscous material is compressed between parallel plates at a rate given by:

$$\frac{dh}{dt} = \frac{2\pi}{3\eta} \frac{Wh^5}{V^2}, \tag{3}$$

where η is the viscosity. This equation is of the form of Equation 2 with $b = 5$. In the solder tests, the values of b obtained were in general much larger than 5.0, so that these materials are not strictly viscous. On the other hand, the fact that the results are independent of the initial sample height indicates that the rate of flow is independent of the strain; while the fact that the curves of log dh/dt vs. log h continue to be linear at very low rates of flow indicates the absence of any yield point, or minimum stress required for flow. Hence the stress required for compression appears to be of the type which has been called quasi-viscous—wholly dependent on the velocity gradient.

Flow of this type, of which viscous flow is a special case, has been observed in colloidal solutions tested by the capillary tube method. In such solutions deWaele [6] found the rate of efflux proportional, not to the pressure (as in strictly viscous fluids [7]), but to a power of the pressure. Porter and Rao [8] have shown that this would be the case if it were assumed that the shearing stress (τ) is proportional to a power of the velocity gradient (dv/dx), or $\tau = \eta' \, (dv/dx)^{1/n}$, where η' and n are constants; dv/dx is the velocity gradient normal to the plane in which τ is the tangential stress. Assuming this relation, the case of compression between parallel plates has been shown [3] to be given by:

$$\frac{dh}{dt} = C \frac{W^n h^{\frac{5(n+1)}{2}}}{V^{\frac{3n+1}{2}}}, \tag{4}$$

where C is a constant inversely proportional to η'. The development of Equation 4 employs approximations corresponding to those used in obtaining Equation 3, involving the assumption that the diameter of the sample is large compared with its height.

As a matter of at least theoretical interest a series of runs were made on one solder (No. 2) employing different loads and sample volumes, and the data thus obtained were analyzed with reference to their agreement with Equation 4. This analysis is given in an appendix

[6] A. de Waele, *J. Oil & Color Chem. Assn.*, **6**, 33 (1923).
[7] Bingham, "Fluidity and Plasticity," McGraw-Hill.
[8] Porter and Rao, *Trans. Faraday Soc.*, **23**, 311 (1927).

to this paper, and shows that all runs could be fitted without significant variation by an equation of the form of Equation 1, and that the values of b for these different runs did not differ significantly from one another. On the other hand, the values of k thus obtained did differ significantly from an expression of the form $k = K(W^a/V^c)$, in which K, a, and c were selected to give the best agreement with the data, indicating that Equation 2, as applied to this material, is not strictly correct. The values of a and c thus obtained differed considerably from the values of n and $(3n + 1)/2$ respectively that would be required if Equation 4 applied, n being given by equating b to $5(n + 1)/2$ in accordance with this equation. On the other hand, these differences were not indicated as necessarily significant, so that it appears that Equation 4 applies to these data to a fair approximation, the divergence being perhaps due wholly to the approximations employed in the theoretical development.

Regardless of the accuracy of Equation 4, the considerations discussed above indicate quite clearly that the flow of solders in the wiping range is quasi-viscous (independent of strain, and of a yield point requirement) and hence similar to the flow of many colloidal solutions. This conclusion is in agreement with the fact that at these temperatures solder consists of a solid phase dispersed in a liquid phase. It is furthermore apparent that the quantity $b - 5$ (or of $1 - n$, if use is made of the theory given) is a measure of the departure of the material from a strictly viscous condition.

Evaluation of Flow Constants

With the exception noted above, the runs made with the various solders tested were confined to a single sample size and a single load. In each case values of b were obtained by plotting log dh/dt vs. log h, values of the former quantity being found by plotting tangents as described. To illustrate the reproducibility of points in the plot of log rate vs. log height, Fig. 2 has been prepared from Fig. 1. Curves I and II of Fig. 1 were plotted to different scales of sample height and the tangents drawn. In Fig. 2 are plotted values of log dh/dt against log h from the two curves. It is seen that the line would be drawn in practically the same place if either set of points was taken alone.

In comparing solders it is evident that the sample height after a given time is a rough measure of relative consistency, the softer material showing the lower sample height. But from Equation 1 it is evident that such a comparison may be misleading for materials showing different values of b. This is brought out in Fig. 3, in which are given the height vs. time curves for two different solders. It will be seen that the curves cross each other, and that the consistency of such samples would, if rated on such a basis, depend upon the time selected for the comparison. In comparing two solders tested under the same conditions, consideration must be given to both constants of Equation 1 (k and b). For purposes of qualitative comparison an equivalent procedure is to consider the sample heights at two different times for each sample listed.

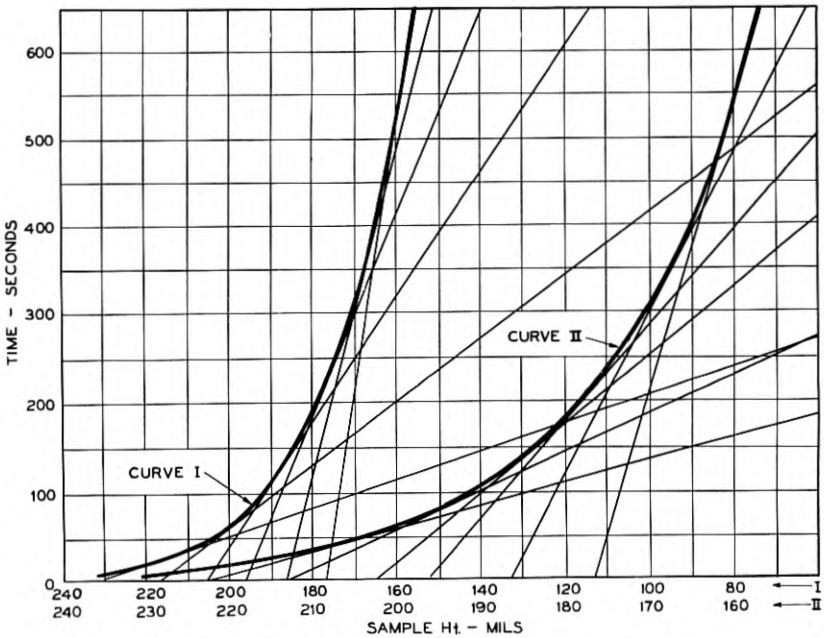


Fig. 1—Plot of a typical set of plasticity data. Solder No. 5—157° C.

Comparison of Plasticity with Solid Liquid Ratio

A summary of the results of the plasticity tests is given in Table II. In this are included values of the percentage of liquid present in the solder at the test temperature, calculated from the equilibrium diagrams by the method advocated by Tammann.[9]

Sample heights after 30 minutes of compression are plotted in Fig. 4 against per cent liquid phase present under the conditions of test. As one would expect, the points for any given solder can be placed on a smooth curve, in which an increase in the proportion of liquid phase

[9] Gustav Tammann, "A Textbook of Metallography."

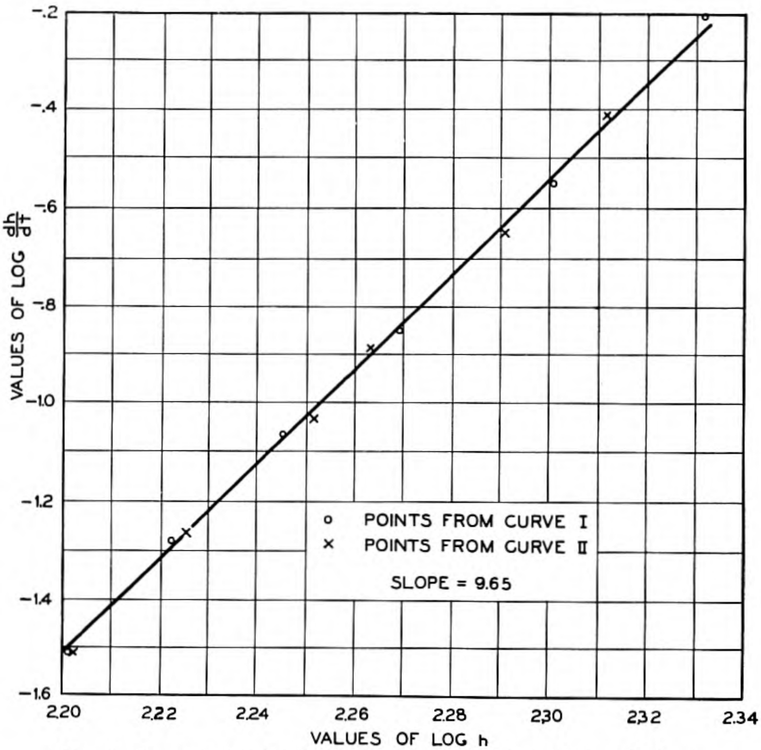


Fig. 2—Log rate vs. log height curve plotted from data of Fig. 1.

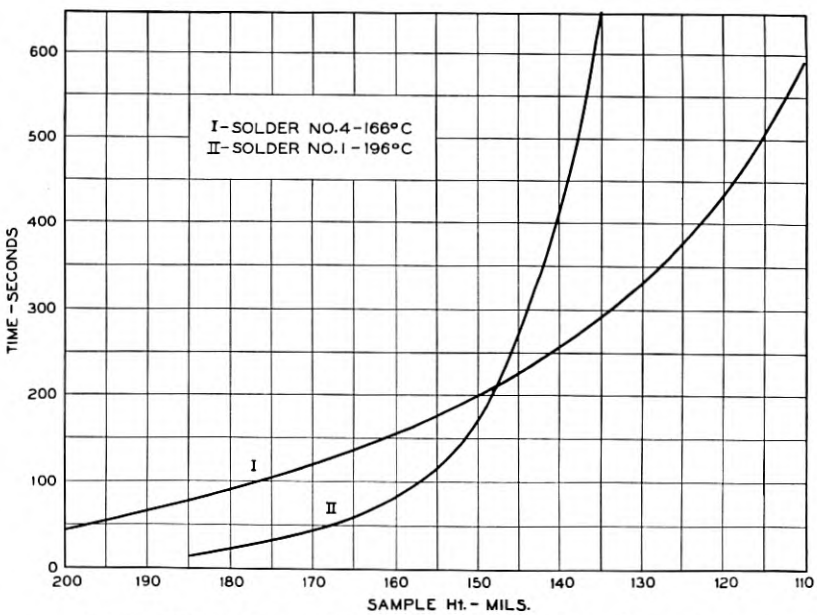


Fig. 3—Effect of type of flow upon sample height vs. time curves.

TABLE II

Results of Plasticity Tests

Solder	Temper-ature °C.	Sample Height 1 Min.	Sample Height 10 Min.	Sample Height 30 Min.	Per Cent Liquid *	*b*	Aver-age *b*
No. 1	196	162	132		47.5	14.58	
	204	145	125	116	51.7	19.12	16.0
	210	117	104	94	55.1	14.20	
No. 2	183	150	128	118	52.4	15.80	
	185	143	124	114	53.4	15.56	16.2
	194	127	111	103	57.8	17.24	
No. 3	170	132	118	110	43.4	25.60	
	184	130	114	105	51.2	18.14	21.0
	192	115	101	93	55.8	19.30	
No. 4	162	266	245	203	55.5	3.88	
	166	197	110	83	59.4	3.58	3.73
	170	42	33	29	62.4	†	
No. 5	149		255			6.02	
	153	245	198	166		6.39	8.24
	157	200	158	140		9.65	
	161	185	154	127		10.90	
No. 6	140	262	244	224		12.3	
	148	214	175	156		10.2	13.5
	156	164	139	127		15.9	
	160	143	124	115		15.6	

* Equilibrium Diagrams used:
Solder No. 1 and 2—Rosenhain & Tucker.
Solder No. 3—Stoffel.
Solder No. 4—International Critical Tables, Vol. IV.
† Sample heights taken were too low to render the data significant.

corresponds to a decrease in the sample height at any specified time. However, the curves for the different solders do not fall together. As a matter of fact, no such agreement would be expected after studying results such as those plotted in Fig. 3. The curves given in this figure show that if a number of solders are tested at temperatures corresponding to a given solid-liquid ratio, they may fall in quite different orders with respect to each other in a comparison of sample heights, according to the time chosen for the comparison.

Still less agreement among the solders is shown when values of b are compared with percentage of liquid phase. There seems to be no relation here; in fact, for a given solder alteration of the temperature and hence of the proportion of liquid phase, does not seem to have much effect upon the value of b.

A survey of the curves and data shows that the possibility of drawing any conclusions from the solid-liquid ratios regarding the plasticity to be expected or regarding the quality of the solder must be dismissed.

Comparison of Workability and Temperature Gradient of Plasticity

In Fig. 5 are shown the plasticity-temperature gradients for the various solders. Here comparative values of plasticity for any one solder are taken as given by the sample heights after ten minutes of compression. It is seen that solders No. 1, 2 and 3 show good (low) gradients, with No. 3 seeming to be superior, while Nos. 5, 6 and especially No. 4, show a large variation in plasticity with temperature. Since, if it is to be properly worked, a wiping solder must not vary rapidly in consistency with temperature, this can be used as one criterion for choosing a satisfactory solder, and on this basis the solders tested should be placed in an order consistent with Fig. 5.

Such a rating is quite in agreement with actual working characteristics, which rank solder No. 3 first, No. 4 last, the others falling in between in essentially the order which Fig. 5 would indicate.

It is now of interest to determine if there exists any relation between the rapidity with which the plasticity varies with temperature and the values of b, the slope of the log rate vs. log height curve. Reference to Table II will show that there is a wide divergence in the values of b found for the different solders. It is further seen that the difference exhibited from one alloy to the next is greater than the difference between individual runs on one solder at different temperatures. This virtual constancy of b for any one solder justifies, it may be noted in passing, the use of sample heights after fixed intervals of compression as comparative measures of plasticity, as in Figs. 4 and 5. There would be no such justification for using such sample heights in comparing different alloys having different values of b.

If then b is regarded as a constant for any one solder, independent of temperature, mean values of b for each solder may be computed from the values obtained at different temperatures. Such mean values have been computed, and are included in Table II. Comparison of these with the curves of Fig. 5 shows a striking agreement between the order of the solders based on increasing temperature gradients and that based on decreasing values of b. This suggests that the temperature gradient of plasticity is the lower, the more removed is the character of the flow from that of a strictly viscous liquid. Furthermore, it is apparent that if the value of b is an inverse measure of the temperature gradient of plasticity, and if the latter is in turn an inverse measure of the workability of a solder, then the value of b is a direct measure of workability.

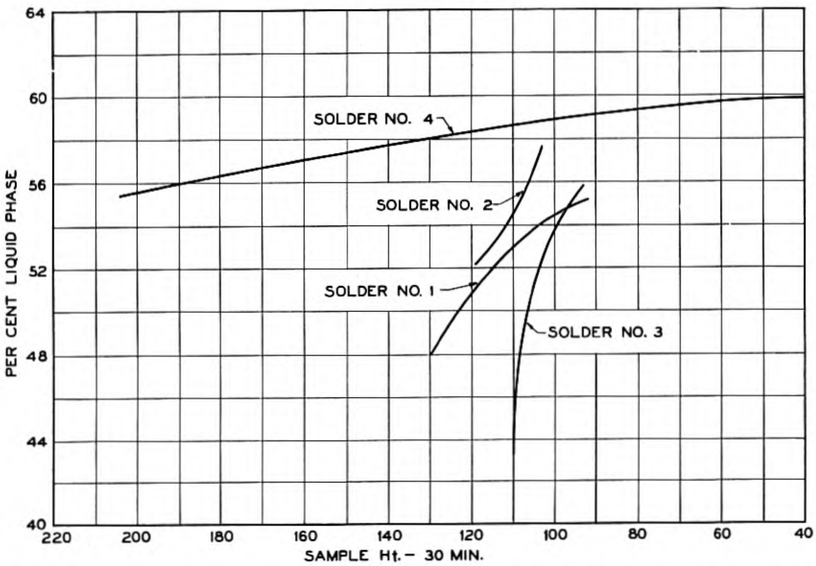


Fig. 4—Relation of plasticity to per cent liquid phase in four of the solders tested.

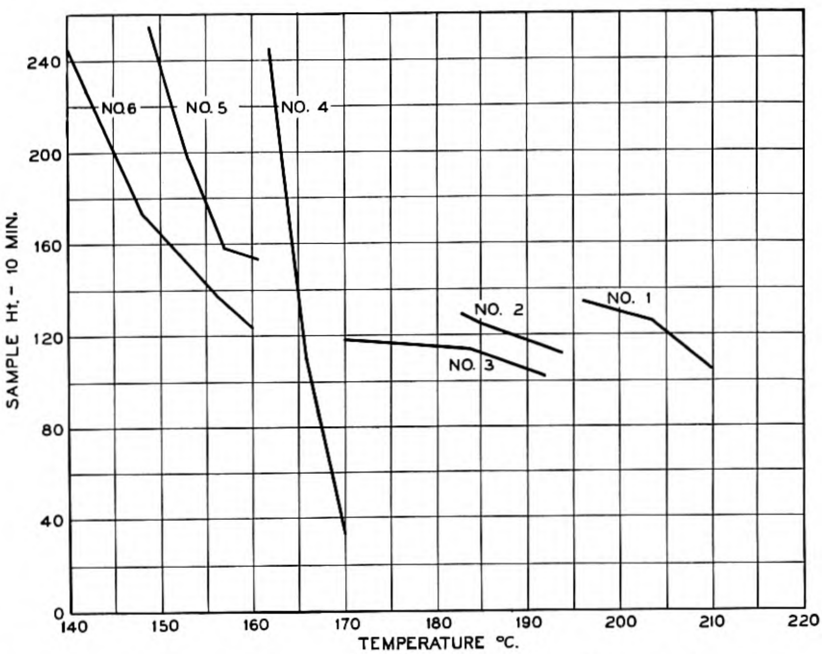


Fig. 5—Variation of plasticity with temperature, showing the superiority of solders Nos. 1, 2 and 3.

As the value of b can be determined in a single test with the plastometer, its relation to workability should permit of the rapid evaluation of the latter property, provided the relation holds in general. Applying this criterion to the solders tested, No. 4, showing virtually a viscous type of flow, ($b = 5$), would be at once thrown out. Solder No. 3 would be classed as a superior solder from the standpoint of workability. Nos. 1 and 2 could be classed as satisfactory. The use of Nos. 5 and 6 would be questionable, with No. 6 having the best chance for success. Practical tests on these last two solders are insufficient to completely confirm this classification, but general indications are that it is correct. Practice has completely confirmed the classification of the other four.

Usefulness of Plastometer in Research and Control Testing

It is believed that the work which has been done on these six solders has furnished information which justifies the use of the plastometer in the future when any group of alloys is to be investigated as to working properties. It appears that the principal demand upon a workable solder is a wide temperature range in which it can be worked. Whether or not a solder is suitable in this respect can very well be determined by the plastometer either by making runs at several temperatures or by determining b at some temperature representative of the wiping range.

A further possible use of the plastometer is in testing solders manufactured or purchased subject to requirements as to their composition. As a rule, the sample height after a given time of compression is quite different for alloys differing in composition. This is illustrated by the curves of Fig. 6, which represent the data obtained with the plastometer on a number of alloys. These show that such a test of composition would not be infallible in a wide application, as the quite dissimilar alloys C and F give quite similar curves. In distinguishing an excess or deficiency of one component in any given series of alloys, however, the test should be quite sensitive.

If the plastometer were used for this purpose, the allowable limits of composition would correspond to limits of sample height in satisfactory samples. Consider, for example, the case of a lead-tin solder of a composition between that of solders Nos. 1 and 2. A convenient temperature of test would be 195° C. At this temperature, the sample height after ten minutes would be 110 mils for a solder of composition 65 per cent Pb, 35 per cent Sn, and 137 mils for one of composition 60 per cent Pb, 40 per cent Sn. Since the plastometer results are reproducible within 2 mils, the accuracy in detecting variation in composition should be easily within 0.5 per cent. Whether the possible usefulness of the test could be realized economically in engineering practice has not as yet been ascertained.

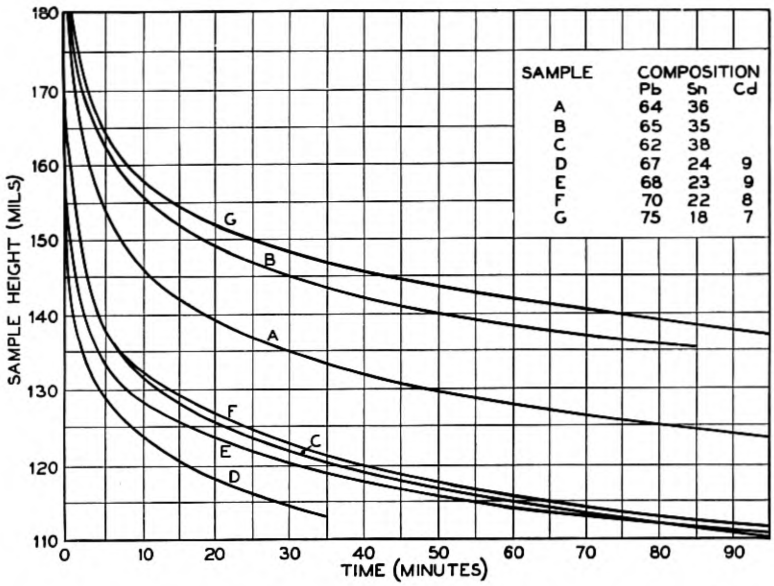


Fig. 6—A group of plasticity curves illustrating the effect of composition upon plasticity.

Segregation

Experimental Determination

The plastic deformation of a material composed of solid and liquid phases, such as a solder during wiping, presents a condition favorable to segregation, and this phenomenon has often been thought to play an important role in determining the quality of a solder. Especially has this been true in regard to porosity. Now, if a sample of solder which has a tendency to segregate is subjected to a plasticity test, a considerable amount of liquid phase should be pushed to the outer edges of the sample during the first part of the run. Since the liquid phase always contains that part of the sample which later solidifies as the eutectic, it follows that after the sample has been allowed to cool, the outer edges should be richer in eutectic than the central portion.

This, as a matter of fact, was found to be true, and measurements of segregation were made upon a number of samples taken from the plastometer. These samples were usually about 5 cm. in diameter. They were cut diametrically, polished and etched.

Photographs were taken with a magnification of 25 ×, which allowed 0.5 cm. of the sample to be photographed at one time. Three photographs were taken along the radius of each sample, one bordered by the axis of the sample, one by the edge, and one mid-way between these two.

These photomicrographs were placed in a beam of light of constant intensity. Since the eutectic appears as a light background, while the particles which separate out as solid above the eutectic temperature appear dark, the light transmitted through the plate is a comparative measure of the relative amount of eutectic present in different parts of the sample. The variation in light transmitted as the picture is passed before the beam therefore indicates the amount of segregation. The transmitted light was directed on a cæsium photo-cell and the amplified photo-electric current measured, four measurements being required to span each plate, making twelve measurements on each sample.

Table III gives the results of these measurements. In this table, only the averages of the four measurements of each photomicrograph are given.

TABLE III

SEGREGATION MEASUREMENTS

Light Transmitted Measured in Milliamperes of Amplified Photo-electric Current

Sample		Ligh Transmitted—Ma.		
Solder	Temp. ° C.	.25 cn . from Center	1.25 cm. from Center	2.25 cm. from Center
1.	204	0.36	1.08	2.00
	210	0.35	1.07	1.55
2.	183	0.88	1.58	3.20
	186	0.44	0.69	1.16
	194	0.22	0.69	0.81
3.	170	0.45	0.57	1.06
	184	0.83	0.87	0.99
	194	0.49	0.57	0.69
4.	162	0.63	0.49	0.58
	166	0.66	0.63	0.63
	170	0.61	0.60	0.65

Results—Comparison with Porosity

The only conclusion that can be drawn from this tabulation of results is that segregation is not an important factor in porosity. Among solders Nos. 1, 2 and 3, good and poor solders alike show segregation to a large extent. No marked difference seems to exist, for example, between the segregation of solder No. 2, which forms non-porous joints, and solder No. 1, which is a poor solder from the standpoint of porosity. On the other hand, in solder No. 4, an extremely poor solder, segregation is so slight as to be negligible. In Fig. 7 are plotted the data for solders Nos. 1 and 4, the former showing

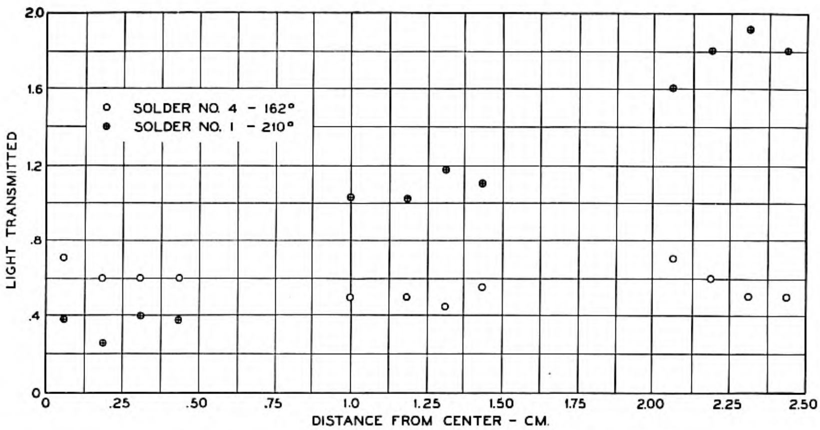


Fig. 7—Segregation measurements on two solder samples.

considerable segregation and the latter practically no segregation of solid from liquid phase.

It is interesting to note how well the segregation measurements confirm the plasticity data. Solders Nos. 1, 2, and 3, showing clearly the quasi-viscous type of flow, also show marked segregation, revealing the reluctance of the solid particles to move and their consequent piling up near the center of the sample. In solder No. 4, the flow is of the viscous type, and consequently there is no segregation. The solid phase and the liquid phase do not separate, but flow together as a viscous whole.

Particle Size

Relation to Porosity

It was thought that in solders forming porous joints the solid phase may, at wiping temperatures, be present as particles relatively large in size, in which case the liquid must be in larger recesses than would be the case if the particles were small. The cohesion of the solder while it is being worked and while it is solidifying must depend to some extent upon the adhesion between solid and liquid. The larger the particles, the greater the distances through which the surface forces must operate in hindering the free motion of the liquid phase. The liquid phase in a two-phase system where the particles are large should then be only loosely bound to the mass, and though there might be no gross tendency toward segregation, this liquid phase should in working be squeezed out, leading to cavity formation and consequent porosity.

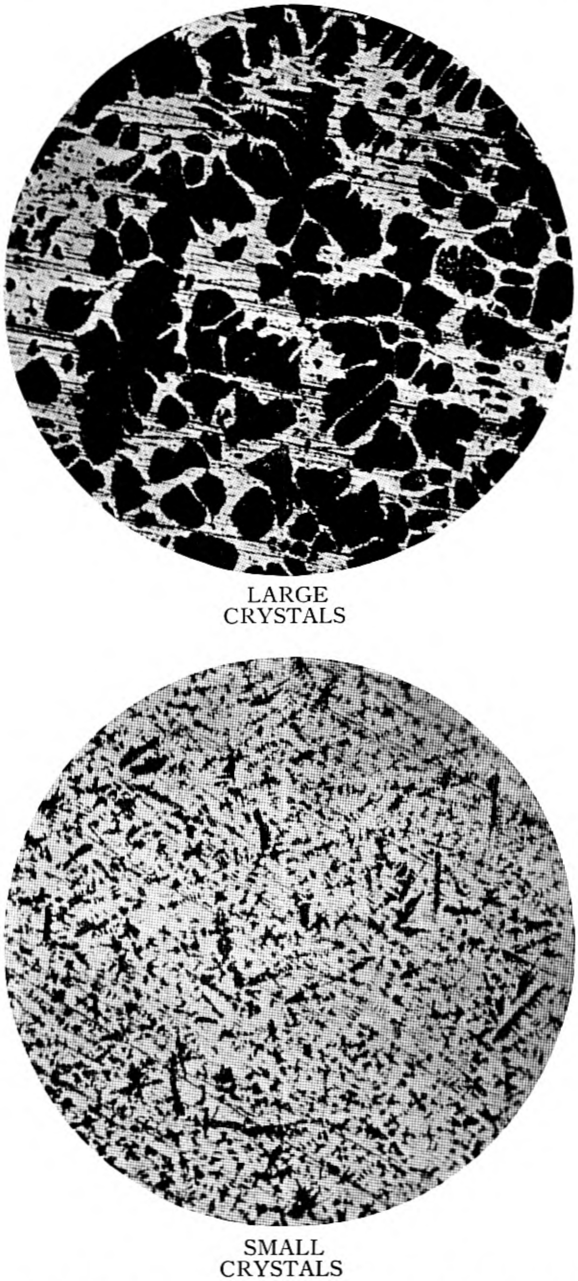
Experimental

To test this explanation of porosity two samples of solder No. 1 were quenched from 210° C. and two of solder No. 2 from 200° C., at which temperatures these solders exhibit practically the same plasticity. Though the data are too meagre to allow generalizations to be made, the results in both sets of samples were well in line with the theory. Microscopic examination showed a much coarser particle structure in solder No. 1 than in solder No. 2, as illustrated by Fig. 8. In this instance at least the solder which forms porous joints tended to form solid particles at wiping temperatures larger than those formed in the solder which is acceptable from the standpoint of porosity.

Particle Size and Porosity as Related to Wettability

It appears that this test may have a significance beyond that just attached to it, and since it may conceivably be used again in the study of solders, it is perhaps well to point out a further explanation based on the theory of interfacial tension.

As was stated earlier in this paper, one demand upon a wiping solder is that it shall wet the cable covering readily. This is necessary if porosity is to be avoided, since areas of imperfect wetting will form tracks between the solder and the lead allowing passage of air or liquid along the interface. Further, the liquid phase of the solder must wet the solid phase of the solder; otherwise passages will exist in the body of the joint. There is some question as to whether porosity in a joint is more commonly due to lack of adherence between the sheath and the solder or whether complete air passages may exist throughout the solder itself. At any rate, there is no assurance that either type of failure does not exist. These two types of porosity are, however, very similar phenomena from the standpoint of wettability, since both the cable sheath and the solid particles of the solder are composed chiefly of lead. If the liquid phase of the solder will wet the sheath, it will also wet the solid particles of lead within the solder, and neither type of failure will be likely to occur. The problem resolves itself into one of the wettability of lead by the liquid phase of the alloy. Therefore, if it is possible to determine, at equal plasticities, the relative



POROUS
WIPING SOLDER
65% Pb. – 35% Sn.
AIR COOLED TO 210° C.
AND QUENCHED

LARGE
CRYSTALS

NON — POROUS
WIPING SOLDER
60% Pb. – 40% Sn.
AIR COOLED TO 200° C.
AND QUENCHED

SMALL
CRYSTALS

Fig. 8—The porosity of wiping solders as related to the crystal size of samples air cooled to equivalent consistencies and quenched.

adherence of liquid to solid for different solders, the relative likelihood of porous joints occurring should be thereby indicated.

Wettability depends upon interfacial tension. As pointed out by Osterhof and Bartell,[10] high wetting power of a liquid for a solid means a low interfacial tension between the two, which in turn means a high adhesion tension.

Furthermore, if the interfacial tension between a liquid and a dispersed solid is high, the formation of large grains, crystals, or flocculates will take place, thus diminishing the amount of solid-liquid interface and the total free energy of the system. This has been used to account for the increased solubility of small particles in contact with a saturated solution, with the consequent growth of large crystals [11, 12] and in paint technology to explain the formation of groups of particles or flocculates.[13, 14] If, on the other hand, the interfacial tension is low, as for solders of high wettability, the particles have a tendency to remain small. This fact has been used to account for the high development of surface in colloidal solutions.

This would lead to the conclusions that in solder No. 1, where large solid particles exist, the liquid phase wets the lead dispersed within it and hence also the lead of the cable sheath but poorly, while in solder No. 2, exhibiting finer grain formation, the wetting power of the liquid phase for lead is high. These considerations alone would point to a conclusion which has been shown to be the case—that solder No. 1 is more likely to form porous joints than solder No. 2.

Summary

An attempt has been made to correlate important qualities of wiping solders with measurements of plasticity, segregation, and particle size.

A study of six solders has shown that their deformation in the working range is in accord with the theory of quasi-viscous flow, which assumes the shearing stress to be proportional to a fractional power of the velocity gradient.

It appears that what is required of a solder, in order that it can be properly worked, is a low variation in plasticity with temperature. This allows the solder to be worked for a comparatively long period while it is losing heat to the atmosphere. This requirement appears to be satisfied for solders whose resistance to deformation differs most markedly in character from that of a viscous liquid.

[10] H. J. Osterhof and F. E. Bartell, *J. Phys. Chem.*, **34**, 7 (1930).
[11] H. Freundlich, "Colloid and Capillary Chemistry," E. P. Dutton (1922).
[12] Willows and Hatschek, "Surface Tension and Surface Energy," P. Blakiston & Son (1919).
[13] Bartell and Van Loo, *Ind. & Eng. Chem.*, **17**, 1051 (1925).
[14] Wm. Green, *Ind. & Eng. Chem.*, **15**, 122 (1923).

Determinations of the rate of compression of solders at wiping temperatures between parallel plates have been found to conform to the equation:

$$\frac{dh}{dt} = kh^b,$$

in which dh/dt is the rate of compression, h the sample height, and k and b are constants, the latter depending only on the character of the material. For viscous materials $b = 5.0$; the larger the value of b, apparently, the more does the flow differ from that of a viscous liquid. Solders having large values of b (15–25) have low temperature gradients of plasticity and good workability, in agreement with the conclusions noted above.

It has been shown that the plasticity cannot be predicted from the solid-liquid ratio.

Evidence is presented that segregation is not responsible for the porosity of wiped joints, and that this defect may be related to the particle size of the solid phase at wiping temperatures. The relation of particle size to cavity formation and to the cohesion and wetting power of the solder is discussed.

Appendix

As stated in the body of this paper, a series of runs were made on solder No. 2 at 183° C. to determine how closely the results of such tests agree with the theoretical expression for quasi-viscous flow given as Equation (4) above. Only five test runs were made, the minimum sufficient to indicate the existence of such agreement. The sample volumes and loads used in these runs are listed in Table IV. For each

TABLE IV

Run No.	V c.c.	W kgs.	b	σ_b	log k	S	log k'	S'	Δ log k'
1.	5.45	13.6	16.10	4.32	−34.36	0.0384	−34.472	0.0352	0.073
2.	7.34	13.6	17.73	3.99	−37.58	0.0351	−35.556	0.0473	0.122
3.	11.01	13.6	15.68	1.85	−35.53	0.0225	−36.577	0.0257	0.067
4.	7.30	20.8	15.76	1.43	−34.38	0.0131	−35.209	0.01553	0.098
5.	7.33	27.2	15.51	2.83	−33.51	0.0249	−34.853	0.0286	0.078

$$\bar{b} = 16.156$$
$$\sigma_L' = 0.906$$
$$a = 1.656$$
$$c = 6.875$$
$$\log K = -31.359$$

run there was prepared a plot of sample height (h) vs. time (t), and values of dh/dt were determined from the slopes of tangents drawn as shown in Fig. 1. There were thus obtained for each run a set of six or seven pairs of values of log dh/dt vs. log h.

The first question to be determined is whether for any run these pairs of values agree with a relation of the form of Equation 1, which can be written:

$$\log \frac{dh}{dt} = \log k + b \log h. \tag{5}$$

As sample height readings appear to be reproducible in such tests to within about one per cent, the standard deviation of values of log h about their correct values may be estimated as rather less than 0.004 (log 1.01). The time readings being more accurate, values of log dh/dt should vary with a standard deviation of like magnitude. If the data can be fitted by an equation of the form of Equation 5, estimates of k and b can be obtained by minimizing the sum of the squares of the deviations of the experimental points from the straight line, these deviations being measured along normals to the line. As for these data b is large, these normal deviations are nearly equal to their horizontal components, the deviations of the observed from the calculated values of log h, and hence the sum of the squares of the latter were minimized. Writing X_i for values of log h and Y_i for corresponding values of log dh/dt, the following computations were carried out for each run (n being the number of pairs of values employed in each case):

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad \overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n},$$

$$\sigma_x^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}, \qquad \sigma_y^2 = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n},$$

$$b = \frac{n\,\sigma_y^2}{\sum_{i=1}^{n} X_i Y_i - n\,\overline{X}\,\overline{Y}}, \qquad \log k = \overline{Y} - b\,\overline{X},$$

$$S^2 = \frac{\sum_{i=1}^{n} (Y_i - \log k - b\,X_i)^2}{n-2}.$$

Values of b, log k, and S thus determined are included in Table IV. The quantity S is an estimate of the standard deviation of the observed from the calculated values of log dh/dt evaluated as recommended by Fisher,[15] page 118. To determine in each case whether these deviations are significant, they may be compared with those that might be expected on the basis of the estimated precision of measurement. As the value of b is approximately 16 in each case, an error of 0.004 in log h corresponds to an error of 0.064 in log dh/dt. The values of S computed are all smaller than this amount, indicating that the differences between observed values of log dh/dt and those calculated from Equation 1 are due to experimental variations.

The next question to be determined is whether the differences observed in the values of b given by the different runs are significant. In Table IV is included the mean value $\bar{b}$ of the values of b, together with the estimated standard deviation of the individual values of b from their true mean, taken as given by:

$$\sigma_b' = \sqrt{\frac{\sum(b - \bar{b})^2}{n - 1}}.$$

This estimate of the standard deviation of the actual values of b may be compared with estimates of the standard deviation of b made for each run on the basis of the variability of the data. An expression for such an estimate is given by Fisher (loc. cit.) as:

$$\sigma_b^2 = \frac{nS^2}{\sigma_x^2}.$$

Values of σ_b as given by this last expression are included in Table IV. These are larger than the value (0.906) of σ_b', computed as described, and it is therefore evident that the data are consistent with the hypothesis that b is constant throughout the series of runs.

To obtain fairer estimates of log k for the subsequent computation, the mean value of b, $\bar{b}$, was taken as giving the slope of the straight line of Equation 5. On this basis the values of log k are given by:

$$\log k' = \bar{Y} - \bar{b}\bar{X}.$$

While the values of S are given by:

$$S'^2 = \frac{\sum_{i=1}^{n} (Y_i - \log k' - \bar{b}X_i)^2}{n - 1}.$$

Values of log k' and S' are given in Table IV. The next step in the computation is to determine if the values of log k thus determined

[15] R. A. Fisher, "Statistical Methods for Research Workers," London (1925).

vary significantly from those given by Equation 2 above, or by:

$$\log k' = \log K + a \log W - c \log V. \qquad (6)$$

For this purpose there were determined values of log K, a and c for which the sum of the squares of the differences of observed and calculated values of log k' are a minimum, the components of deviation due to variations in values of log W and log V being neglected, as these are known to a higher order of accuracy than are values of log k'. Following the usual procedure for calculating partial regression coefficients, the following computations were performed, writing Z_i, Y_i, and X_i for corresponding values of log k', log W and log V:

$$\bar{Z} = \frac{\sum_{i=1}^{n} Z_i}{n}, \qquad \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}, \qquad \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n},$$

$$\sigma_Z^2 = \frac{\sum_{i=1}^{n} (Z_i - \bar{Z})^2}{n}, \qquad \sigma_Y^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n}, \qquad \sigma_X^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n},$$

$$\sum xy = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}, \qquad \sum yz = \sum_{i=1}^{n} Y_i Z_i - n\bar{Y}\bar{Z},$$

$$\sum zx = \sum_{i=1}^{n} Z_i X_i - n\bar{Z}\bar{X},$$

$$a = \frac{n\sigma_x^2 \sum yz - \sum xy \cdot \sum xz}{n^2\sigma_x^2\sigma_y^2 - (\sum xy)^2}, \qquad -c = \frac{n\sigma_y^2 \sum xz - \sum xy \cdot \sum yz}{n^2\sigma_x^2\sigma_y^2 - (\sum xy)^2},$$

$$\log K = \bar{Z} - a\bar{Y} + c\bar{X}.$$

The values of a, c, and log K thus obtained are included in Table IV, together with values headed Δ log k', which are the differences between the values of log k' listed in Table IV and those calculated from Equation 6, using the listed values a, c, and log K. To determine whether these differences are significant or not, the probability of obtaining a value in error to that extent may be calculated for each run from the value (S') of the standard deviation of log k' estimated by the method given by Fisher (loc. cit.). These probabilities are evaluated from Table IV of the text cited, taking n as $n' - 3$, where n' is the number of pairs of values of log dh/dt and log h used to evaluate log k', and taking t as:

$$t = \frac{\Delta \log k' \sqrt{n'}}{S'}.$$

The values of t thus computed range from 5.4 to 16.7 and correspond in every case to a probability less than 0.01, so that not one of the actual values of log k' should be experimentally in error to an extent sufficient to account for its deviation from the value given by Equation 6. Hence Equation 2, is at best only an approximation to a correct expression for dh/dt.

Of course, Equation 2 is actually quite a fair approximation to an expression for dh/dt, as is evident in comparing the differences Δ log k' with the differences among the values of log k' themselves. In fact, if values of log k' are plotted against log W for a constant value of V, or against log V for a constant value of W, quite good straight lines are obtained, and the inaccuracy of the approximation can only be brought out by an analysis such as is here given. Whether this inaccuracy is due to the approximations employed in the theoretical development of Equation 4, or to an error in the basic assumption that $\tau = [dv/dx]^{1/n}$, cannot, of course, be determined. It is, however, of interest to see if the values of a and c are related to that of b in the manner required by Equation 4. For $b = 16.156$ (the value of $\bar{b}$ observed), Equation 4 requires that $a = 5.462$ and $c = 8.693$. The observed values of a and c are 1.656 and 6.875 respectively. It remains to determine if these differences may be accidental, provided the values of log k' differ from those given by Equation 6 as greatly as observed. Following Fisher (page 133), the probability of a divergence between observed and calculated values of c of $c_1 - c_2$ is given by entering Table IV of the text with $n = n' - 3$, where n' is the number of cases (5), and with t given by:

$$t = \frac{C_1 - C_2}{\sqrt{\dfrac{[Z - (\bar{Z} + a\bar{Y} - c\bar{X})]^2}{n-3} \cdot \dfrac{n\sigma_y^2}{n^2\sigma_x^2\sigma_y^2 - (\sum xy)^2}}} = 3.58.$$

This corresponds to a probability between 0.05 and 0.10, which indicates that the divergence may be due to chance, though it is more likely to represent a real difference. Similarly the probability of obtaining the observed difference $a_1 - a_2$ is given from the table by entering with $n = 2$ and with t given by:

$$t = \frac{a_1 - a_2}{\sqrt{\dfrac{[Z - (\bar{Z} + a\bar{Y} - c\bar{X})]^2}{n-3} \cdot \dfrac{n\sigma_x^2}{n^2\sigma_x^2\sigma_y^2 - (\sum xy)^2}}} = 5.85.$$

This corresponds to a probability between 0.02 and 0.05, which suggests more strongly than does the other case that the divergence is real. For so few observations, however, the distribution theory on which Fisher's method is based cannot give a definite indication when the probabilities lie so close to the arbitrary dividing line (0.05) between likely and unlikely results. The analysis therefore leaves open the question as to whether the theoretical relation given by Equation 4 between a, b and c is valid within the limits of variability imposed by the approximate nature of any relation of the form of Equation 2. That any equation of this form is only approximately in agreement with these data is conclusively shown by the above analysis.

Regeneration Theory

By H. NYQUIST

Regeneration or feed-back is of considerable importance in many applications of vacuum tubes. The most obvious example is that of vacuum tube oscillators, where the feed-back is carried beyond the singing point. Another application is the 21-circuit test of balance, in which the current due to the unbalance between two impedances is fed back, the gain being increased until singing occurs. Still other applications are cases where portions of the output current of amplifiers are fed back to the input either unintentionally or by design. For the purpose of investigating the stability of such devices they may be looked on as amplifiers whose output is connected to the input through a transducer. This paper deals with the theory of stability of such systems.

Preliminary Discussion

WHEN the output of an amplifier is connected to the input through a transducer the resulting combination may be either stable or unstable. The circuit will be said to be stable when an impressed small disturbance, which itself dies out, results in a response which dies out. It will be said to be unstable when such a disturbance results in a response which goes on indefinitely, either staying at a relatively small value or increasing until it is limited by the non-linearity of the amplifier. When thus limited, the disturbance does not grow further. The net gain of the round trip circuit is then zero. Otherwise stated, the more the response increases the more does the non-linearity decrease the gain until at the point of operation the gain of the amplifier is just equal to the loss in the feed-back admittance. An oscillator under these conditions would ordinarily be called stable but it will simplify the present paper to use the definitions above and call it unstable. Now, this fact as to equality of gain and loss appears to be an accident connected with the non-linearity of the circuit and far from throwing light on the conditions for stability actually diverts attention from the essential facts. In the present discussion this difficulty will be avoided by the use of a strictly linear amplifier, which implies an amplifier of unlimited power carrying capacity. The attention will then be centered on whether an initial impulse dies out or results in a runaway condition. If a runaway condition takes place in such an amplifier, it follows that a non-linear amplifier having the same gain for small current and decreasing gain with increasing current will be unstable as well.

Steady-State Theories and Experience

First, a discussion will be made of certain steady-state theories; and reasons why they are unsatisfactory will be pointed out. The most obvious method may be referred to as the series treatment. Let the complex quantity $AJ(i\omega)$ represent the ratio by which the amplifier and feed-back circuit modify the current in one round trip, that is, let the magnitude of AJ represent the ratio numerically and let the angle of AJ represent the phase shift. It will be convenient to refer to AJ as an admittance, although it does not have the dimensions of the quantity usually so called. Let the current

$$I_0 = \cos \omega t = \text{real part of } e^{i\omega t} \qquad (a)$$

be impressed on the circuit. The first round trip is then represented by

$$I_1 = \text{real part of } AJe^{i\omega t} \qquad (b)$$

and the nth by

$$I_m = \text{real part of } A^nJ^ne^{i\omega t}. \qquad (c)$$

The total current of the original impressed current and the first n round trips is

$$I_n = \text{real part of } (1 + AJ + A^2J^2 + \cdots A^nJ^n)e^{i\omega t}. \qquad (d)$$

If the expression in parentheses converges as n increases indefinitely, the conclusion is that the total current equals the limit of (d) as n increases indefinitely. Now

$$1 + AJ + \cdots A^nJ^n = \frac{1 - A^{n+1}J^{n+1}}{1 - AJ}\cdot \qquad (e)$$

If $|AJ| < 1$ this converges to $1/(1 - AJ)$ which leads to an answer which accords with experiment. When $|AJ| > 1$ an examination of the numerator in (e) shows that the expression does not converge but can be made as great as desired by taking n sufficiently large. The most obvious conclusion is that when $|AJ| > 1$ for some frequency there is a runaway condition. This disagrees with experiment, for instance, in the case where AJ is a negative quantity numerically greater than one. The next suggestion is to assume that somehow the expression $1/(1 - AJ)$ may be used instead of the limit of (e). This, however, in addition to being arbitrary, disagrees with experimental results in the case where AJ is positive and greater than 1, where the expression $1/(1 - AJ)$ leads to a finite current but where experiment indicates an unstable condition.

The fundamental difficulty with this method can be made apparent by considering the nature of the current expressed by (*a*) above. Does the expression $\cos \omega t$ indicate a current which has been going on for all time or was the current zero up to a certain time and $\cos \omega t$ thereafter? In the former case we introduce infinities into our expressions and make the equations invalid; in the latter case there will be transients or building-up processes whose importance may increase as n increases but which are tacitly neglected in equations (*b*) – (*e*). Briefly then, the difficulty with this method is that it neglects the building-up processes.

Another method is as follows: Let the voltage (or current) at any point be made up of two components

$$V = V_1 + V_2, \tag{f}$$

where V is the total voltage, V_1 is the part due directly to the impressed voltage, that is to say, without the feed-back, and V_2 is the component due to feed-back alone. We have

$$V_2 = AJV. \tag{g}$$

Eliminating V_2 between (*f*) and (*g*)

$$V = V_1/(1 - AJ). \tag{h}$$

This result agrees with experiment when $|AJ| < 1$ but does not generally agree when AJ is positive and greater than unity. The difficulty with this method is that it does not investigate whether or not a steady state exists. It simply assumes tacitly that a steady state exists and if so it gives the correct value. When a steady state does not exist this method yields no information, nor does it give any information as to whether or not a steady state exists, which is the important point.

The experimental facts do not appear to have been formulated precisely but appear to be well known to those working with these circuits. They may be stated loosely as follows: There is an unstable condition whenever there is at least one frequency for which AJ is positive and greater than unity. On the other hand, when AJ is negative it may be very much greater than unity and the condition is nevertheless stable. There are instances of $|AJ|$ being about 100 without the conditions being unstable. This, as will appear, accords closely with the rule deduced below.

Notation and Restrictions

The following notation will be used in connection with integrals:

$$\int_I \phi(z)dz = \lim_{M\to\infty} \int_{-iM}^{+iM} \phi(z)dz, \tag{1}$$

the path of integration being along the imaginary axis (see equation 9), i.e., the straight line joining $-iM$ and $+iM$;

$$\int_{s+} \phi(z)dz = \lim_{M\to\infty} \int_{-iM}^{iM} \phi(z)dz, \tag{2}$$

the path of integration being along a semicircle [1] having the origin for center and passing through the points $-iM$, M, iM;

$$\int_C \phi(z)dz = \lim_{M\to\infty} \int_{-iM}^{-iM} \phi(z)dz, \tag{3}$$

the path of integration being first along the semicircle referred to and then along a straight line from iM to $-iM$. Referring to Fig. 1 it

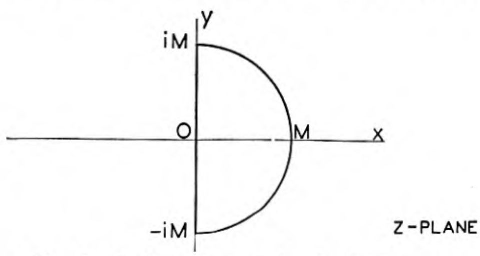


Fig. 1—Paths of integration in the z-plane.

will be seen that

$$\int_{s+} - \int_I = \int_C . \tag{4}$$

The total feed-back circuit is made up of an amplifier in tandem with a network. The amplifier is characterized by the amplifying ratio A which is independent of frequency. The network is characterized by the ratio $J(i\omega)$ which is a function of frequency but does not depend on the gain. The total effect of the amplifier and the network is to multiply the wave by the ratio $AJ(i\omega)$. An alternative way of characterizing the amplifier and network is to say that the amplifier is

[1] For physical interpretation of paths of integration for which $x > 0$ reference is made to a paper by J. R. Carson, "Notes on the Heaviside Operational Calculus," *B. S. T. J.*, Jan. 1930. For purposes of the present discussion the semicircle is preferable to the path there discussed.

characterized by the amplifying factor A which is independent of time, and the network by the real function $G(t)$ which is the response caused by a unit impulse applied at time $t = 0$. The combined effect of the amplifier and network is to convert a unit impulse to the function $AG(t)$. Both these characterizations will be used.

The restrictions which are imposed on the functions in order that the subsequent reasoning may be valid will now be stated. There is no restriction on A other than that it should be real and independent of time and frequency. In stating the restrictions on the network it is convenient to begin with the expression G. They are

$$G(t) \text{ has bounded variation, } -\infty < t < \infty. \tag{AI}$$

$$G(t) = 0, \qquad -\infty < t < 0. \tag{AII}$$

$$\int_{-\infty}^{\infty} |G(t)|\,dt \text{ exists.} \tag{AIII}$$

It may be shown [2] that under these conditions $G(t)$ may be expressed by the equation

$$G(t) = \frac{1}{2\pi i}\int_I J(i\omega)e^{i\omega t}d(i\omega), \tag{5}$$

where

$$J(i\omega) = \int_{-\infty}^{\infty} G(t)e^{-i\omega t}dt. \tag{6}$$

These expressions may be taken to define J. The function may, however, be obtained directly from computations or measurements; in the latter case the function is not defined for negative values of ω. It must be defined as follows to be consistent with the definition in (6):

$$J(-i\omega) = \text{complex conjugate of } J(i\omega). \tag{7}$$

While the final results will be expressed in terms of $AJ(i\omega)$ it will be convenient for the purpose of the intervening mathematics to define an auxiliary and closely related function

$$w(z) = \frac{1}{2\pi i}\int_I \frac{AJ(i\omega)}{i\omega - z}d(i\omega), \qquad 0 < x < \infty, \tag{8}$$

where

$$z = x + iy \tag{9}$$

and where x and y are real. Further, we shall define

$$w(iy) = \lim_{x \to 0} w(z). \tag{10}$$

[2] See Appendix II for fuller discussion.

The function will not be defined for $x < 0$ nor for $|z| = \infty$. As defined it is analytic[3] for $0 < x < \infty$ and at least continuous for $x = 0$.

The following restrictions on the network may be deduced:

$$\lim_{y \to \infty} y|J(iy)| \text{ exists.} \tag{BI}$$

$$J(iy) \text{ is continuous.} \tag{BII}$$

$$w(iy) = AJ(iy). \tag{BIII}$$

Equation (5) may now be written

$$AG(t) = \frac{1}{2\pi i}\int_I w(z)e^{zt}dz = \frac{1}{2\pi i}\int_{s+} w(z)e^{zt}dz. \tag{11}$$

From a physical standpoint these restrictions are not of consequence. Any network made up of positive resistances, conductances, inductances, and capacitances meets them. Restriction (AII) says that the response must not precede the cause and is obviously fulfilled physically. Restriction (AIII) is fulfilled if the response dies out at least exponentially, which is also assured. Restriction (AI) says that the transmission must fall off with frequency. Physically there are always enough distributed constants present to insure this. This effect will be illustrated in example 8 below. Every physical network falls off in transmission sooner or later and it is ample for our purposes if it begins to fall off, say, at optical frequencies. We may say then that the reasoning applies to all linear networks which occur in nature. It also applies to other linear networks which are not physically producible but which may be specified mathematically. See example 7 below.

A temporary wave $f_0(t)$ is to be introduced into the system and an investigation will be made of whether the resultant disturbance in the system dies out. It has associated with it a function $F(z)$ defined by

$$f_0(t) = \frac{1}{2\pi i}\int_I F(z)e^{zt}dz = \frac{1}{2\pi i}\int_{s+} F(z)e^{zt}dz. \tag{12}$$

$F(z)$ and $f_0(t)$ are to be made subject to the same restrictions as $w(z)$ and $G(t)$ respectively.

Derivation of a Series for the Total Current

Let the amplifier be linear and of infinite power-carrying capacity. Let the output be connected to the input in such a way that the

[3] W. F. Osgood, "Lehrbuch der Funktionentheorie," 5th ed., Kap. 7, § 1, Hauptsatz. For definition of "analytic" see Kap. 6, § 5.

amplification ratio for one round trip is equal to the complex quantity AJ, where A is a function of the gain only and J is a function of ω only, being defined for all values of frequency from 0 to ∞.

Let the disturbing wave $f_0(t)$ be applied anywhere in the circuit. We have

$$f_0(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(i\omega)e^{i\omega t}d\omega \tag{13}$$

or

$$f_0(t) = \frac{1}{2\pi i}\int_{s+} F(z)e^{zt}dz. \tag{13'}$$

The wave traverses the circuit and on completing the first trip it becomes

$$f_1(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} w(i\omega)F(i\omega)e^{i\omega t}d\omega \tag{14}$$

$$= \frac{1}{2\pi i}\int_{s+} w(z)F(z)e^{zt}dz. \tag{14'}$$

After traversing the circuit a second time it becomes

$$f_2(t) = \frac{1}{2\pi i}\int_{s+} Fw^2e^{zt}dz, \tag{15}$$

and after traversing the circuit n times

$$f_n(t) = \frac{1}{2\pi i}\int_{s+} Fw^ne^{zt}dz. \tag{16}$$

Adding the voltage of the original impulse and the first n round trips we have a total of

$$s_n(t) = \sum_{k=0}^{n} f_k(t) = \frac{1}{2\pi i}\int_{s+} F(1 + w + \cdots w^n)e^{zt}dz. \tag{17}$$

The total voltage at the point in question at the time t is given by the limiting value which (17) approaches as n is increased indefinitely [4]

$$s(t) = \sum_{k=0}^{\infty} f_k(t) = \lim_{n\to\infty}\frac{1}{2\pi i}\int_{s+} S_n(z)e^{zt}dz, \tag{18}$$

where

$$S_n = F + Fw + Fw^2 + \cdots Fw^n = \frac{F(1 - w^{n+1})}{1 - w}. \tag{19}$$

[4] Mr. Carson has called my attention to the fact that this series can also be derived from Theorem IX, p. 49, of his Electric Circuit Theory. Whereas the present derivation is analogous to the theory expressed in equations (*a*)–(*e*) above, the alternative derivation would be analogous to that in equations (*f*)–(*h*).

Convergence of Series

We shall next prove that the limit $s(t)$ exists for all finite values of t. It may be stated as of incidental interest that the limit

$$\int_{s+} S_\infty(z)e^{izt}dz \tag{20}$$

does not necessarily exist although the limit $s(t)$ does. Choose M_0 and N such that

$$|f_0(\lambda)| \leq M_0. \qquad 0 \leq \lambda \leq t. \tag{21}$$

$$|G(t-\lambda)| \leq N. \quad 0 \leq \lambda \leq t. \tag{22}$$

We may write [5]

$$f_1(t) = \int_{-\infty}^{\infty} G(t-\lambda)f_0(\lambda)d\lambda. \tag{23}$$

$$|f_1(t)| \leq \int_0^t M_0 N d\lambda = M_0 N t. \tag{24}$$

$$f_2(t) = \int_{-\infty}^{\infty} G(t-\lambda)f_1(\lambda)d\lambda. \tag{25}$$

$$|f_2(t)| \leq \int_0^t M_0 N^2 t dt = M_0 N^2 t^2/2! \tag{26}$$

Similarly

$$|f_n(t)| \leq M_0 N^n t^n/n! \tag{27}$$

$$|s_n(t)| \leq M_0(1 + Nt + \cdots N^n t^n/n!). \tag{28}$$

It is shown in almost any text [6] dealing with the convergence of series that the series in parentheses converges to e^{Nt} as n increases indefinitely. Consequently, $s_n(t)$ converges absolutely as n increases indefinitely.

Relation Between $s(t)$ and w

Next consider what happens to $s(t)$ as t increases. As t increases indefinitely $s(t)$ may converge to zero, indicating a condition of stability, or it may go beyond any value however large, indicating a runaway condition. The question which presents itself is: *Referring to (18) and (19), what properties of $w(z)$ and further what properties of $AJ(i\omega)$ determine whether $s(t)$ converges to zero or diverges as t increases*

[5] G. A. Campbell, "Fourier Integral," *B. S. T. J.*, Oct. 1928, Pair 202.
[6] E.g., Whittaker and Watson, "Modern Analysis," 2d ed., p. 531.

indefinitely? From (18) and (19)

$$s(t) = \lim_{n \to \infty} \frac{1}{2\pi i} \int_{s+} F\left(\frac{1}{1-w} - \frac{w^{n+1}}{1-w}\right) e^{zt} dz. \tag{29}$$

We may write

$$s(t) = \frac{1}{2\pi i} \int_{s+} [F/(1-w)] e^{zt} dz - \lim_{n \to \infty} \frac{1}{2\pi i} \int_{s+} [F w^{n+1}/(1-w)] e^{zt} dz \tag{30}$$

provided these functions exist. Let them be called $q_0(t)$ and $\lim_{n \to \infty} q_n(t)$ respectively. Then

$$q_n(t) = \int_{-\infty}^{\infty} q_0(t-\lambda) \phi(\lambda) d\lambda. \tag{31}$$

where

$$\phi(\lambda) = \frac{1}{2\pi i} \int_{s+} w^{n+1} e^{z\lambda} dz. \tag{32}$$

By the methods used under the discussion of convergence above it can then be shown that this expression exists and approaches zero as n increases indefinitely provided $q_0(t)$ exists and is equal to zero for $t < 0$. Equation (29) may therefore be written, subject to these conditions

$$s(t) = \frac{1}{2\pi i} \int_{s+} [F/(1-w)] e^{zt} dz. \tag{33}$$

In the first place the integral is zero for negative values of t because the integrand approaches zero faster than the path of integration increases. Moreover,

$$\int_I [F/(1-w)] e^{zt} dz \tag{34}$$

exists for all values of t and approaches zero for large values of t if $1 - w$ does not equal zero on the imaginary axis. Moreover, the integral

$$\int_C [F/(1-w)] e^{zt} dz \tag{35}$$

exists because

1. Since F and w are both analytic within the curve the integrand does not have any essential singularity there,
2. The poles, if any, lie within a finite distance of the origin because $w \to 0$ as $|z|$ increases, and
3. These two statements insure that the total number of poles is finite.

We shall next evaluate the integral for a very large value of t. It will suffice to take the C integral since the I integral approaches zero. Assume originally that $1 - w$ does not have a root on the imaginary axis and that $F(z)$ has the special value $w'(z)$. The integral may be written

$$\frac{1}{2\pi i}\int_C [w'/(1 - w)]e^{zt}dz. \tag{36}$$

Changing variables it becomes

$$\frac{1}{2\pi i}\int_D [1/(1 - w)]e^{zt}dw, \tag{37}$$

where z is a function of w and D is the curve in the w plane which corresponds to the curve C in the z plane. More specifically the imaginary axis becomes the locus $x = 0$ and the semicircle becomes a small curve which spirals around the origin. See Fig. 2. The function

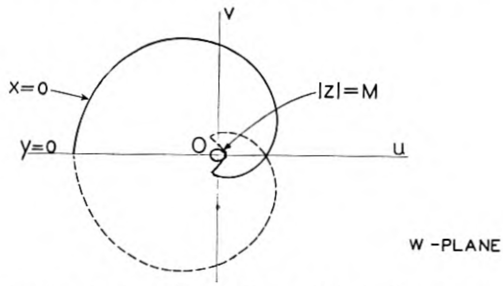


Fig. 2—Representative paths of integration in the w-plane corresponding to paths in Fig. 1.

z and, therefore, the integrand is, in general, multivalued and the curve of integration must be considered as carried out over the appropriate Riemann surface.[7]

Now let the path of integration shrink, taking care that it does not shrink across the pole at $w = 1$ and initially that it does not shrink across such branch points as interfere with its passage, if any. This shrinking does not alter the integral [8] because the integrand is analytic at all other points. At branch points which interfere with the passage of the path the branches stopped may be severed, transposed and connected in such a way that the shrinking may be continued past the branch point. This can be done without altering the value of the integral. Thus the curve can be shrunk until it becomes one or more very small circles surrounding the pole. The value of the total integral

[7] Osgood, loc. cit., Kap. 8.
[8] Osgood, loc. cit., Kap. 7, § 3, Satz 1.

(for very large values of t) is by the method of residues [9]

$$\sum_{j=1}^{n} r_j e^{z_j t}, \tag{38}$$

where z_j ($j = 1, 2 \cdots n$) is a root of $1 - w = 0$ and r_j is its order. The real part of z_j is positive because the curve in Fig. 1 encloses points with $x > 0$ only. The system is therefore stable or unstable according to whether

$$\sum_{j=1}^{n} r_j$$

is equal to zero or not. But the latter expression is seen from the procedure just gone through to equal the number of times that the locus $x = 0$ encircles the point $w = 1$.

If F does not equal w' the calculation is somewhat longer but not essentially different. The integral then equals

$$\sum_{j=1}^{n} \frac{F(z_j)}{w(z_j)} e^{z_j t} \tag{39}$$

if all the roots of $1 - w = 0$ are distinct. If the roots are not distinct the expression becomes

$$\sum_{j=1}^{n} \sum_{k=1}^{r_j} A_{jk} t^{k-1} e^{z_j t}, \tag{40}$$

where A_{jr_j}, at least, is finite and different from zero for general values of F. It appears then that unless F is specially chosen the result is essentially the same as for $F = w'$. The circuit is stable if the point lies wholly outside the locus $x = 0$. It is unstable if the point is within the curve. It can also be shown that if the point is on the curve conditions are unstable. We may now enunciate the following

Rule: Plot plus and minus the imaginary part of $AJ(i\omega)$ against the real part for all frequencies from 0 to ∞. If the point $1 + i0$ lies completely outside this curve the system is stable; if not it is unstable.

In case of doubt as to whether a point is inside or outside the curve the following criterion may be used: Draw a line from the point ($u = 1$, $v = 0$) to the point $z = -i\infty$. Keep one end of the line fixed at ($u = 1$, $v = 0$) and let the other end describe the curve from $z = -i\infty$ to $z = i\infty$, these two points being the same in the w plane. If the net angle through which the line turns is zero the point ($u = 1$, $v = 0$) is on the outside, otherwise it is on the inside.

If AJ be written $|AJ|(\cos\theta + i\sin\theta)$ and if the angle always

[9] Osgood, loc. cit., Kap. 7, § 11, Satz 1.

changes in the same direction with increasing ω, where ω is real, the rule can be stated as follows: The system is stable or unstable according to whether or not a real frequency exists for which the feed-back ratio is real and equal to or greater than unity.

In case $d\theta/d\omega$ changes sign we may have the case illustrated in Figs. 3 and 4. In these cases there are frequencies for which w is real and

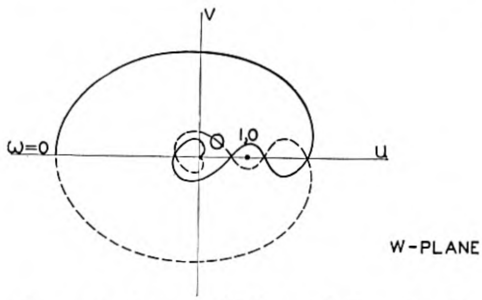


Fig. 3—Illustrating case where amplifying ratio is real and greater than unity for two frequencies, but where nevertheless the path of integration does not include the point 1, 0.

greater than 1. On the other hand, the point (1, 0) is outside of the locus $x = 0$ and, therefore, according to the rule there is a stable condition.

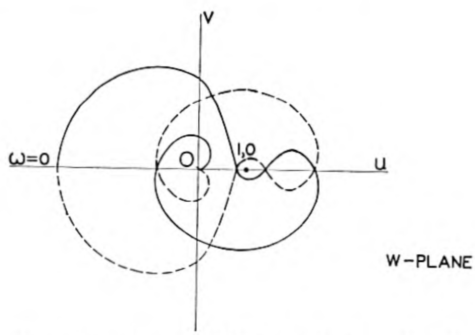


Fig. 4—Illustrating case where amplifying ratio is real and greater than unity for two frequencies, but where nevertheless the path of integration does not include the point 1, 0.

If networks of this type were used we should have the following interesting sequence of events: For low values of A the system is in a stable condition. Then as the gain is increased gradually, the system becomes unstable. Then as the gain is increased gradually still further, the system again becomes stable. As the gain is still further increased the system may again become unstable.

Examples

The following examples are intended to give a more detailed picture of certain rather simple special cases. They serve to illustrate the previous discussion. In all the cases F is taken equal to AJ so that f_0 is equal to AG. This simplifies the discussion but does not detract from the illustrative value.

1. Let the network be pure resistance except for the distortionless amplifier and a single bridged condenser, and let the amplifier be such that there is no reversal. We have

$$AJ(i\omega) = \frac{B}{\alpha + i\omega}, \tag{41}$$

where A and α are real positive constants. In (18) [10]

$$\begin{aligned} f_n &= \frac{1}{2\pi i}\int_I A^{n+1}J^{n+1}(i\omega)e^{i\omega t}di\omega \\ &= Be^{-\alpha t}(B^n t^n/n!). \end{aligned} \tag{42}$$

$$s(t) = Be^{-\alpha t}(1 + Bt + B^2t^2/2! + \cdots). \tag{43}$$

The successive terms f_0, f_1, etc., represent the impressed wave and the successive round trips. The whole series is the total current.

It is suggested that the reader should sketch the first few terms graphically for $B = \alpha$, and sketch the admittance diagrams for $B < \alpha$, and $B > \alpha$.

The expression in parentheses equals e^{Bt} and

$$s(t) = Be^{(B-\alpha)t}. \tag{44}$$

This expression will be seen to converge to 0 as t increases or fail to do so according to whether $B < \alpha$ or $B \geqq \alpha$. This will be found to check the rule as applied to the admittance diagram.

2. Let the network be as in 1 except that the amplifier is so arranged that there is a reversal. Then

$$AJ(i\omega) = \frac{-B}{\alpha + i\omega}\cdot \tag{45}$$

$$f_n = (-1)^{n+1}Be^{-\alpha t}(B^n t^n/n!). \tag{46}$$

The solution is the same as in 1 except that every other term in the series has its sign reversed:

$$\begin{aligned} s(t) &= -Be^{-\alpha t}(1 - Bt + B^2t^2/2! + \cdots) \\ &= -Be^{(-\alpha-B)t}. \end{aligned} \tag{47}$$

[10] Campbell, loc. cit. Pair 105.

This converges to 0 as t increases regardless of how great B may be taken. If the admittance diagram is drawn this is again found to check the rule.

3. Let the network be as in 1 except that there are two separated condensers bridged across resistance circuits. Then

$$AJ(i\omega) = \frac{B^2}{(\alpha + i\omega)^2} \cdot \tag{48}$$

The solution for $s(t)$ is obtained most simply by taking every other term in the series obtained in 1.

$$\begin{aligned} s(t) &= Be^{-\alpha t}(Bt + B^3t^3/3! + \cdots) \\ &= Be^{-\alpha t} \sinh Bt. \end{aligned} \tag{49}$$

4. Let the network be as in 3 except that there is a reversal. Then

$$AJ(i\omega) = \frac{-B^2}{(\alpha + i\omega)^2} \cdot \tag{50}$$

The solution is obtained most directly by reversing the sign of every other term in the series obtained in 3.

$$\begin{aligned} s(t) &= -Be^{-\alpha t}(Bt - B^3t^3/3! + \cdots) \\ &= -Be^{-\alpha t} \sin Bt. \end{aligned} \tag{51}$$

This is a most instructive example. An approximate diagram has been made in Fig. 5, which shows that as the gain is increased the

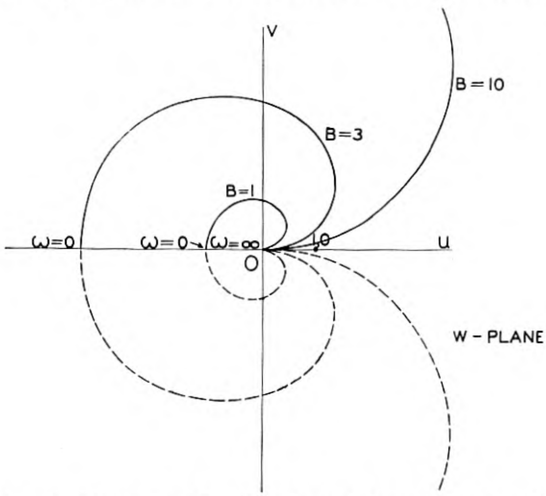


Fig. 5—Illustrating Example 4, with three values for B.

feed-back ratio may be made arbitrarily great and the angle arbitrarily small without the condition being unstable. This agrees with the expression just obtained, which shows that the only effect of increasing the gain is to increase the frequency of the resulting transient.

5. Let the conditions be as in 1 and 3 except for the fact that four separated condensers are used. Then

$$AJ(i\omega) = \frac{B^4}{(\alpha + i\omega)^4} \cdot \tag{52}$$

The solution is most readily obtained by selecting every fourth term in the series obtained in 1.

$$\begin{aligned} s(t) &= Be^{-\alpha t}(B^3t^3/3! + B^7t^7/7! + \cdots) \\ &= \tfrac{1}{2}Be^{-\alpha t}\,(\sinh Bt - \sin Bt). \end{aligned} \tag{53}$$

This indicates a condition of instability when $B \geqq \alpha$, agreeing with the result deducible from the admittance diagram.

6. Let the conditions be as in 5 except that there is a reversal. Then

$$Y = \frac{-\,B^4}{(\alpha + i\omega)^4} \cdot \tag{54}$$

The solution is most readily obtained by changing the sign of every other term in the series obtained in 5.

$$s(t) = Be^{-\alpha t}(-\,B^3t^3/3! + B^7t^7/7! - \cdots). \tag{55}$$

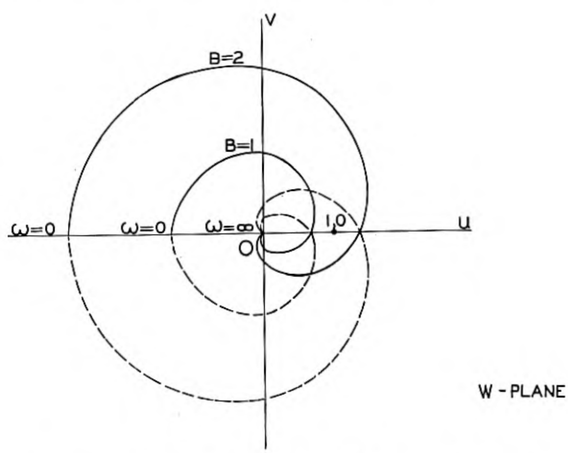


Fig. 6—Illustrating Example 6, with two values for B.

For large values of t this approaches

$$s(t) = -\tfrac{1}{2}Be^{(B/\sqrt{2}-\alpha)t}\sin(Bt/\sqrt{2} - \pi/4). \quad (56)$$

This example is interesting because it shows a case of instability although there is a reversal. Fig. 6 shows the admittance diagram for $B\sqrt{2} - \alpha < 0$ and for $B\sqrt{2} - \alpha > 0$.

7. Let

$$AG(t) = f_0(t) = A(1 - t), \qquad 0 \leq t \leq 1. \quad (57)$$

$$AG(t) = f_0(t) = 0, \qquad -\infty < t < 0, \qquad 1 < t < \infty. \quad (57')$$

We have

$$AJ(i\omega) = A\int_0^1 (1 - t)e^{-i\omega t}dt$$

$$= A\left(\frac{1 - e^{-i\omega}}{\omega^2} + \frac{1}{i\omega}\right). \quad (58)$$

Fig. 7 is a plot of this case for $A = 1$.

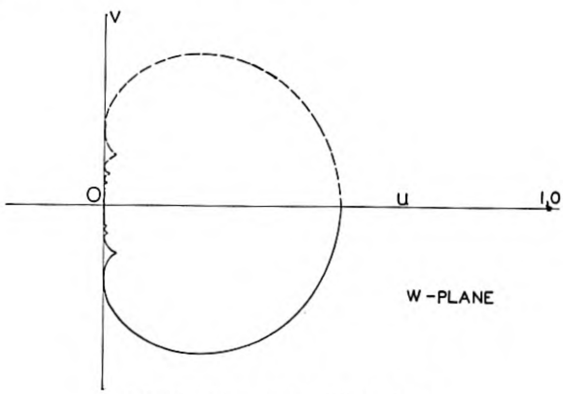


Fig. 7—Illustrating Example 7.

8. Let

$$AJ(i\omega) = \frac{A(1 + i\omega)}{(1 + i2\omega)}. \quad (59)$$

This is plotted on Fig. 8 for $A = 3$. It will be seen that the point 1 lies outside of the locus and for that reason we should expect that the system would be stable. We should expect from inspecting the diagram that the system would be stable for $A < 1$ and $A > 2$ and that it would be unstable for $1 \leq A \leq 2$. We have overlooked one fact, however; the expression for $AJ(i\omega)$ does not approach zero as ω

increases indefinitely. Therefore, it does not come within restriction (BI) and consequently the reasoning leading up to the rule does not apply.

The admittance in question can be made up by bridging a capacity in series with a resistance across a resistance line. This admittance

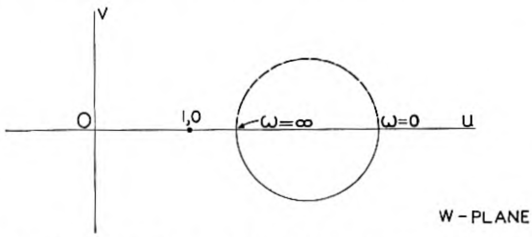


Fig. 8—Illustrating Example 8, without distributed constants.

obviously does not approach zero as the frequency increases. In any actual network there would, however, be a small amount of distributed capacity which, as the frequency is increased indefinitely, would cause the transmission through the network to approach zero. This is shown graphically in Fig. 9. The effect of the distributed capacity is

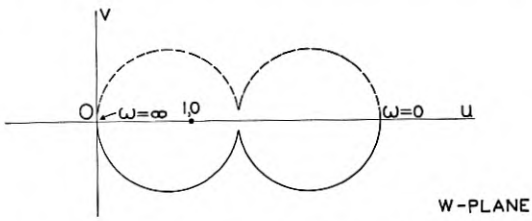


Fig. 9—Illustrating Example 8, with distributed constants.

essentially to cut a corridor from the circle in Fig. 8 to the origin, which insures that the point lies inside the locus.

Appendix I

Alternative Procedure

In some cases $AJ(i\omega)$ may be given as an analytic expression in $(i\omega)$. In that case the analytic expression may be used to define w for all values of z for which it exists. If the value for $AJ(i\omega)$ satisfies all the restrictions the value thus defined equals the w defined above for $0 \leq x < \infty$ only. For $-\infty < x < 0$ it equals the analytic continuation of the function w defined above. If there are no essential singularities anywhere including at ∞, the integral in (33) may be evaluated by the theory of residues by completing the path of integration so that all the poles of the integrand are included. We then have

$$s(t) = \sum_{j=1}^{j=n} \sum_{k=1}^{r_j} A_{jk} t^{k-1} e^{z_j t}. \tag{60}$$

If the network is made up of a finite number of lumped constants there is no essential singularity and the preceding expression converges because it has only a finite number of terms. In other cases there is an infinite number of terms, but the expression may still be expected to converge, at least, in the usual case. Then the system is stable if all the roots of $1 - w = 0$ have $x < 0$. If some of the roots have $x \geqq 0$ the system is unstable.

The calculation then divides into three parts:

1. The recognition that the impedance function is $1 - w$.[11]

2. The determination of whether the impedance function has zeros for which $x \geqq 0$.[12]

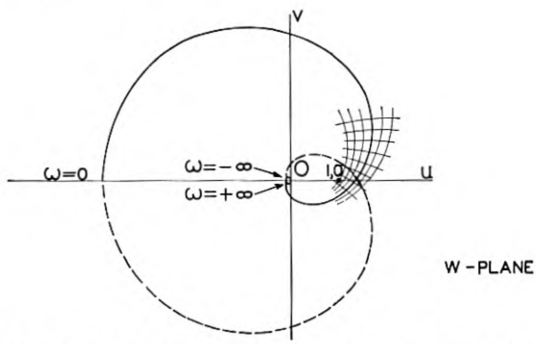


Fig. 10—Network of loci x = const., and y = const.

3. A deduction of a rule for determining whether there are roots for which $x \geqq 0$. The actual solution of the equation is usually too laborious.

To proceed with the third step, plot the locus $x = 0$ in the w plane, i.e., plot the imaginary part of w against the real part for all the values of y, $-\infty < y < \infty$. See Fig. 10. Other loci representing

$$x = \text{const.} \tag{61}$$

and

$$y = \text{const.} \tag{62}$$

[11] Cf. H. W. Nichols, *Phys. Rev.*, vol. 10, pp. 171–193, 1917.
[12] Cf. Thompson and Tait, "Natural Philosophy," vol. I, § 344.

may be considered and are indicated by the network shown in the figure in fine lines. On one side of the curve x is positive and on the other it is negative. Consider the equation

$$w(z) - 1 = 0$$

and what happens to it as A increases from a very small to a very large value. At first the locus $x = 0$ lies wholly to the left of the point. For this case the roots must have $x < 0$. As A increases there may come a time when the curve or successive convolutions of it will sweep over the point $w = 1$. For every such crossing at least one of the roots changes the sign of its x. We conclude that if the point $w = 1$ lies inside the curve the system is unstable. It is now possible to enunciate the rule as given in the main part of the paper but there deduced with what appears to be a more general method.

Appendix II

Discussion of Restrictions

The purpose of this appendix is to discuss more fully the restrictions which are placed on the functions defining the network. A full discussion in the main text would have interrupted the main argument too much.

Define an additional function

$$n(z) = \frac{1}{2\pi i}\int_I \frac{AJ(i\lambda)}{i\lambda - z}\,d(i\lambda), \qquad -\infty < x < 0. \tag{63}$$

$$n(iy) = \lim_{x\to 0} n(z).$$

This definition is similar to that for $w(z)$ given previously. It is shown in the theorem [13] referred to that these functions are analytic for $x \neq 0$ if $AJ(i\omega)$ is continuous. We have not proved, as yet, that the restrictions placed on $G(t)$ necessarily imply that $J(i\omega)$ is continuous. For the time being we shall assume that $J(i\omega)$ may have finite discontinuities. The theorem need not be restricted to the case where $J(i\omega)$ is continuous. From an examination of the second proof it will be seen to be sufficient that $\int_I J(i\omega)d(i\omega)$ exist. Moreover, that proof can be slightly modified to include all cases where conditions (AI)–(AIII) are satisfied.

[13] Osgood, loc. cit.

For, from the equation at top of page 298 [13]

$$\left|\frac{w(z_0 - \Delta z) - w(z_0)}{\Delta z} - \frac{1}{2\pi i}\int_I \frac{AJ(i\lambda)}{(i\lambda - z_0)^2}d(i\lambda)\right|$$

$$\leq |\Delta z|\left|\frac{1}{2\pi i}\int_I \frac{AJ(i\lambda)d(i\lambda)}{(i\lambda - z_0 - \Delta z)(i\lambda - z_0)^2}\right|, \qquad x_0 > 0. \quad (64)$$

It is required to show that the integral exists. Now

$$\int_I \frac{AJ(i\lambda)d(i\lambda)}{(i\lambda - z_0 - \Delta z)(i\lambda - z_0)^2}$$

$$= \int_I \frac{AJ(i\lambda)d(i\lambda)}{(i\lambda - z_0)^3}\left(1 + \frac{\Delta z}{i\lambda - z_0} + \frac{\Delta z^2}{i\lambda - z_0} + \text{etc.}\right) \quad (65)$$

if Δz is taken small enough so the series converges. It will be sufficient to confine attention to the first term. Divide the path of integration into three parts,

$$-\infty < \lambda < -|z_0| - 1, \qquad -|z_0| - 1 < \lambda < |z_0| + 1, \qquad |z_0| + 1 < \lambda < \infty.$$

In the middle part the integral exists because both the integrand and the range of integration are finite. In the other ranges the integral exists if the integrand falls off sufficiently rapidly with increasing λ. It is sufficient for this purpose that condition (BI) be satisfied. The same proof applies to $n(z)$.

Next, consider $\lim_{x \to 0} w(z) = w(iy)$. If iy is a point where $J(iy)$ is continuous, a straightforward calculation yields

$$w(iy) = AJ(iy)/2 + P(iy). \quad (66a)$$

Likewise,

$$n(iy) = -AJ(iy)/2 + P(iy) \quad (66b)$$

where $P(iy)$ is the principal value [14] of the integral

$$\frac{1}{2\pi i}\int_I \frac{AJ(i\lambda)}{i\lambda - iy}d(i\lambda).$$

Subtracting

$$w(iy) - n(iy) = AJ(iy) \quad (67)$$

If (iy) is a point of discontinuity of $J(iy)$

$$|w| \text{ and } |n| \text{ increase indefinitely as } x \to 0. \quad (68)$$

Next, evaluate the integral

$$\frac{1}{2\pi i}\int_{x+I} w(z)e^{zt}dz,$$

[14] E. W. Hobson, "Functions of a Real Variable," vol. I, 3d edition, § 352.

where the path of integration is from $x - i\infty$ to $x + i\infty$ along the line $x =$ const. On account of the analytic nature of the integrand this integral is independent of x (for $x > 0$). It may be written then

$$\lim_{x\to 0}\frac{1}{2\pi i}\int_{x+I} w(z)e^{zt}dz = \lim_{x\to 0}\frac{1}{2\pi i}\int_{x+I}\frac{1}{2\pi i}\int_I \frac{AJ(i\lambda)}{i\lambda - z}e^{zt}d(i\lambda)dz$$

$$= \lim_{x\to 0}\frac{1}{2\pi i}\int_{x+I}\frac{1}{2\pi i}\lim_{M\to\infty}\left[\int_{-iM}^{iy-i\delta} + \int_{iy-i\delta}^{iy+i\delta} + \int_{iy+i\delta}^{iM}\right]\frac{AJ(i\lambda)}{i\lambda - z}e^{zt}d(i\lambda)dz$$

$$= \lim_{x\to 0}\left[\frac{1}{2\pi i}\int_{x+I}\frac{1}{2\pi i}\int_{iy-i\delta}^{iy+i\delta}\frac{AJ(i\lambda)}{i\lambda - z}e^{zt}d(i\lambda)dz + Q(t,\delta)\right],\ x>0, \quad (69)$$

where δ is real and positive. The function Q defined by this equation exists for all values of t and for all values of δ. Similarly,

$$\lim_{x\to 0}\frac{1}{2\pi i}\int_{x+I} n(z)e^{zt}dz$$

$$= \left[\lim_{x\to 0}\frac{1}{2\pi i}\int_{x+I}\frac{1}{2\pi i}\int_{iy+i\delta}^{iy+i\delta}\frac{AJ(i\lambda)}{i\lambda - z}e^{zt}d(i\lambda)dz + Q(t,\delta)\right],\ x<0, \quad (70)$$

Subtracting and dropping the limit designations

$$\frac{1}{2\pi i}\int_{x+I} w(z)e^{zt}dz - \frac{1}{2\pi i}\int_{x+I} n(z)e^{zt}dz = \frac{1}{2\pi i}\int_I AJ(i\lambda)e^{i\lambda t}d(i\lambda). \quad (71)$$

The first integral is zero for $t < 0$ as can be seen by taking x sufficiently large. Likewise, the second is equal to zero for $t > 0$. Therefore,

$$\frac{1}{2\pi i}\int_{x+I} w(z)e^{zt}dz = \frac{1}{2\pi i}\int_I AJ(i\omega)e^{i\omega t}d(i\omega) = AG(t), \quad 0 < t < \infty \quad (72)$$

$$-\frac{1}{2\pi i}\int_{x+I} n(z)e^{zt}dz$$

$$= \frac{1}{2\pi i}\int_I AJ(i\omega)e^{i\omega t}d(i\omega) = AG(t) \quad -\infty < t < 0. \quad (73)$$

We may now conclude that

$$\int_I n(iy)e^{iyt}d(iy) = 0, \qquad -\infty < t < \infty \quad (74)$$

provided

$$G(t) = 0, \qquad -\infty < t < 0. \quad \text{(AII)}$$

But (74) is equivalent to

$$n(z) = 0, \quad (74')$$

which taken with (67) gives

$$w(iy) = AJ(iy). \tag{BIII}$$

(BIII) is, therefore, a necessary consequence of (AII). (74′) taken with (68) shows that

$$J(iy) \text{ is continuous.} \tag{BII}$$

It may be shown [15] that (BI) is a consequence of (AI). Consequently all the B conditions are deducible from the A conditions.

Conversely, it may be inquired whether the A conditions are deducible from the B conditions. This is of interest if $AJ(i\omega)$ is given and is known to satisfy the B conditions, whereas nothing is known about G.

Condition AII is a consequence of BIII as may be seen from (67) and (74). On the other hand AI and AIII cannot be inferred from the B conditions. It can be shown by examining (5), however, that if the slightly more severe condition

$$\lim_{y\to\infty} y^\gamma J(iy) \text{ exists,} \qquad (\gamma > 1), \tag{BIa}$$

is satisfied then

$$G(t) \text{ exists,} \qquad -\infty < t < \infty, \tag{AIa}$$

which, together with AII, insures the validity of the reasoning.

It remains to show that the measured value of $J(i\omega)$ is equal to that defined by (6). The measurement consists essentially in applying a sinusoidal wave and determining the response after a long period. Let the impressed wave be

$$E = \text{real part of } e^{i\omega t}, \qquad t \geq 0. \tag{75}$$
$$E = 0, \qquad t < 0. \tag{75′}$$

The response is

$$\text{real part of} \int_0^t AG(\lambda)e^{i\omega(t-\lambda)}\,d\lambda$$
$$= \text{real part of } Ae^{i\omega t}\int_0^t G(\lambda)e^{-i\omega\lambda}d\lambda. \tag{76}$$

For large values of t this approaches

$$\text{real part of } Ae^{i\omega t}J(i\omega). \tag{77}$$

Consequently, the measurements yield the value $AJ(i\omega)$.

[15] See Hobson, loc. cit., vol. II, 2d edition, § 335. It will be apparent that K depends on the total variation but is independent of the limits of integration.

Contemporary Advances in Physics, XXIII

Data and Nature of Cosmic Rays

By KARL K. DARROW

"Cosmic rays" is the name of the ultimate cause which maintains that part of the ionization of the air which cannot be ascribed to the rays of radioactive substances on earth nor to any other known agency. The measurement of this residue, the discrimination between it and that part of the ionization which is due to familiar rays, is the first problem of cosmic-ray research. Second comes the problem of learning, from measurements made at as many places and under as many conditions as possible, the nature of the mysterious ionizing agent. One naturally begins by assuming it to be like in kind to one or another of the known types of ionizing rays, but different in quality; *e.g.* to consist of electrons faster than any known electrons, or photons of greater energy and lesser wave-length than any known photons. It is not yet certain whether one of these hypotheses will fit, or rays of some new type must be imagined. The direct measurements of ionization are supplemented by observations of material particles of evidently enormous energy which dart across the atmosphere in straight paths leaving trains of ions behind. On the whole it seems highly probable that these particles, or those (if such there be) from which they receive their energy, come to the earth from outer space; and the energy which they bear is so great, that its source must be some process not yet known.

THE subject of this article is unique in modern physics for the minuteness of the phenomena, the delicacy of the observations, the adventurous excursions of the observers, the subtlety of the analysis, and the grandeur of the inferences. The effect which is studied, which may be described as the liberation of electrons from the molecules of the air by agents otherwise unknown, amounts at sea-level to the liberation of only about 1500 of these per cubic decimetre of air per second. It is not the whole of the observed effect; these 1500 electrons are those which are left over out of a quantity often much greater, after allowance is made for the actions of all known ionizing agents. The methods employed are ranked among the most ingenious and sensitive of science; yet the apparatus is not invariably set up in the calm seclusion of the laboratory. Physicists with their frail machines have gone to high mountain ponds in the Sierras and the Andes, to the distant wildernesses about the earth's magnetic poles; they have scooped out cavities in Alpine glaciers, they have lifted hundredweights of lead to the tops of peaks above the snow-line, they have cruised the arctic and the tropical oceans, they have descended into tunnels and deep mines, they have ascended into the sky in aeroplanes and balloons. As for the analysis of the precious data obtained with so much labor, it will be evident from this article how intricate a process it now proves to be.

And the incentive for all these labors? At first, the problem seems little more exciting than the tracing of an insignificant leak in an apparatus supposedly insulated. There is, however, reason to suspect that *this* leak is due to causes really sensational. The comprehended part of the ionization of the air being due to known rays such as X-rays, light, and the radiations from radioactive substances, it is natural to ascribe the mysterious residue to rays as yet unknown. It turns out then that these hypothetical rays must lie beyond any yet discovered; if they are electrons, they must be swifter electrons—if protons, faster protons—if corpuscles of light, then corpuscles of higher frequency and higher energy, than any thus far known; or else they must be particles of a totally new variety. It is not easy to imagine where they could come from on earth, and there are various reasons for supposing that they wander in from outer space. Messengers from the depths of space and from the stars—corpuscles of visible light, meteorites, the electrons which are the presumptive causes of the aurora—are constantly being received and have been amply interpreted; it is probable that these cosmic rays have a message of value, perhaps of the first importance.

The dates of the beginnings of cosmic-ray research have been much disputed, in an unprofitable fashion. I like to set aside the controversies by choosing a time extravagantly remote, the year 1785! In that year, it is said, Coulomb made the first acceptable proof that the air of the atmosphere is conductive. It had long been known that an electrified body, mounted on the best available insulator and set up in the air, slowly loses its charge; it had, however, been thought that the "leak" is partly an escape through the insulator or over the surface thereof, partly a carrying-off of charge by particles of dust which drift up to the electrified body, touch it and depart—Coulomb found that in addition to these, there is an actual conduction through the air.

After this discovery, progress ceased for many years. Towards the end of the nineteenth century it was, however, learned that the air becomes much more conductive than normally it is, when rays of certain kinds pervade it—X-rays for instance, and the radiations from radioactive atoms. Further, it was found that some part, at least, of the normal conductivity of atmospheric air is due to rays from radioactive atoms blown about by the wind or embedded in the ground. Then came the time for asking: Is there a residual part of the normal conductivity of the air, which cannot be ascribed to rays of any kind hitherto known? With the asking of this question, the way was opened to the sequence of researches, which form the topic of this article. "Cosmic rays" are *by definition* the cause of that residue of the ionization of the atmosphere, which is left over after deduction of all of the ionization which can be traced to the action of the rays of radioactive substances, or any other known cause whatever.

I write "ionization" instead of "conductivity"; this is because it is well enough known that when a volume of gas is and remains conductive, it is constantly being ionized—which is to say, electrons are constantly being detached from its molecules. Molecule and electron become ions, the members of an "ion-pair"; if there is an electrified body nearby, those of the sign opposite to its charge are drawn to it (those of the other sign going to the earth) and step by step their charges neutralize its own. The question set above may therefore be rephrased: Are ion-pairs appearing in the atmosphere oftener than rays of known varieties are making them?

To answer this question, one must of course exclude all of the rays from radioactive substances (and, naturally, X-rays and other artificial kinds) from the volume of air which one is studying; or if they cannot be excluded altogether, one must estimate the ionizing-power of those which remain. They are important in cosmic-ray research as causes of error; they are important also in another way, as tentative models for theories of the cause of the residual ionization. In attempting to explain this mysterious residue, the procedure both natural and wise is to begin by supposing it similar to alpha-rays, or else to beta-rays, or else to gamma-rays. In the end it may be found needful to postulate some ionizing-agent entirely different from all of these; but in the beginning, they should be the guides.

Of the three named classes of rays, one (the alpha-particles or positive corpuscles) has a sharply-limited range in air so short, that such as come from the ground are stopped within a few inches; and a range in metal so extremely short, that even if one wished it would not be possible to build a durable air-chamber with solid walls so thin that the alpha-rays could get through.

The second (the beta-particles or fast electrons) have a range which varies from one radioactive element to another, and even from one particle to another among those emitted by a single element. This range is often much longer than that of an alpha-particle, and less precise. Air at a distance of many metres above the ground may be sensibly ionized by beta-corpuscles proceeding from the soil. All along their paths, beta-particles engender ion-pairs profusely—in air at normal (sea-level) density, never fewer than forty per cm. Nevertheless they do not enter an air-chamber walled by a very few centimeters of lead, and need not be a serious cause of confusion.

More formidable is the third kind, the gamma-rays or short electromagnetic waves. They also vary from one element to another, and different wave-lengths may proceed from a single element; some are innocuous, but some are much more penetrating than even beta-rays. One must be careful about the language here; one must not say that they have a great range. To speak of "range" implies that when a number of particles is started off side by side with equal speed along parallel paths into a stratum of matter, they continue to follow parallel paths—albeit with diminishing speed—until they are brought to a stop, this termination coming for each of them at the same distance from their source. Of alpha-particles this is nearly a true statement, of beta-particles it is a fair approximation. Of gamma-corpuscles or photons (for gamma-rays are high-frequency light, and may therefore be regarded alternatively as corpuscles and as waves) it is not true at all. If a number of these, all of identical wave-length, is started off side by side along parallel paths into a stratum of matter, one after another impinges on an atom, or (it is probably better to say) upon an electron belonging to some atom. Until such an impact, the photon does not ionize at all, but passes unperceived. The number which at a distance x from the source have not yet suffered an impact varies as an exponential function $e^{-\mu x}$ of that distance.

At such an impact the gamma-ray photon loses energy and suffers a deflection, which may or may not be considerable. It also liberates the electron, which is an act of ionization; but usually the electron itself, being started off with considerable speed, liberates a great number of additional electrons from other atoms of the gas, so that the direct ionization by the gamma-rays falls far behind the indirect; *most of the ionization by gamma-rays is done by the intermediacy of fast electrons.* One may say that the lower the value of μ, the more penetrating, or "harder," are the gamma-rays; one may compare values of μ for gamma-rays of different wave-lengths; but one may not compare a value of μ for these with a value of range for beta-particles or alpha-particles (unless one should arbitrarily define the range of a beam of gamma-rays as the distance over which its intensity is reduced by say 99 per cent, which would be very misleading).

These are very important matters, for the chance of deciding whether the cosmic rays are likest to gamma-rays or to material particles turns largely upon differences such as these. Another detail is important: the practice of stating values of μ for standard materials, such as water or lead. It is *approximately* true that slabs of various materials will absorb the same fraction of an incident beam of gamma-rays (of a single wave-length, be it understood) if their thicknesses are so adjusted that all of them have the same mass per unit of surface-area. The densities of air at sea-level, of water and of lead stand to each other in about the ratios .0013 : 1 : 11.4; a metre of water is therefore "equivalent" in this respect to about nine centimetres of lead or three-quarters of a kilometre of air at sea-level. After measuring μ in a material of density ρ, one may reduce it to a value approximately equal to that which would be found in a standard material (lead, for instance) of density ρ_0, by multiplying it with the factor (ρ_0/ρ). A better approximation yet, for the harder gamma-rays, is the statement that the slabs absorb the same fraction of the gamma-ray beam if the atoms under each unit of surface-area have the same number of bound electrons altogether. Denoting the number of bound electrons per unit volume of the material in question by E, the corresponding number for lead (say) by E_0, we find (E_0/E) for the value of the factor by which an observed value of μ is to be multiplied, in order to convert it into one approximately valid for lead.[1] To give a definite example, I will mention simply that ten centimetres of lead intercept all but a few per cent of the hardest gamma-ray corpuscles which come from radioactive bodies; and that a couple of kilometres above the earth, one need not worry about the influence of the radioactivity of the ground.[2]

So, to avoid ionization of air by the rays of the known radioactive elements at the surface of the earth, one surrounds the air which one is observing by matter more than thick enough to stop all of the alpha and all of the beta-particles, and thick enough to cut off all but a few per cent of the gamma-corpuscles. Often this is done by using plates of lead several cm. thick for the walls of the air-chamber. Sometimes it is done by sinking the chamber into a lake, or digging a hole into the ice of a glacier. A much greater thickness of water frozen or liquid is required to intercept the unwanted rays, than of lead; this thickness must surround the air-chamber on all sides, towards the bottom of the lake, towards its edges, even (if there be overhanging mountains, or radioactivity in the air) towards its surface; but water is inexpensive, and does not have to be carried about. Or by going up in a balloon, or sending up the air-chamber in a balloon without oneself attending it, one may put enough air beneath the apparatus to screen away the rays from the ground.

But are there not radioactive atoms in the very materials used to

[1] For an element of atomic weight A and atomic number Z and density ρ, the value of E is $Z\rho/Am_H$, the symbol m_H standing for the mass of the hydrogen atom. For lead $Z = 82$, $A = 206$, $\rho_0 = 11.4$. For a compound of known constitution the formula can easily be worked out; if the constitution is unknown but certainly involves atoms of high atomic number only, one may put $Z/A = \frac{1}{2}$ as a rough approximation.

[2] Millikan in a specific case says that 4.8 cm. of lead intercept 90 per cent of the gamma-rays from igneous rocks (*Phys. Rev.* (2), **28**, p. 862; 1926).

protect the ionization-chamber? the air, the ice, the water, the leaden walls themselves? Of the atmospheric air I have already admitted this; the quantity of radioactive gas commingled with it can, however, be measured, is generally very small, and does not seem to bother observers when they are not too close to the ground. As for the water: Millikan has always carefully chosen "snow-fed" lakes, high in the mountains where the water is derived not from springs which have seeped through soil, but from snow which has fallen onto bare rock from on high. Ice of glaciers, it is to be presumed, has an origin equally uncorrupted. The same cannot be said for the water of Lake Constance wherein Regener's data were taken, but he by special tests proved the effect of its radioactivity to be slight.

As for the lead (or whatever other metal or metals may be used in making the ionization-chamber) this too may be contaminated with radioactive atoms. Some people have mounted wire netting all around the interior of the chamber at some distance from the walls, so that such rays as come from the wall and are soft (in the sense opposite to that in which I spoke of gamma-rays as hard) may be largely absorbed between the wall and the net; voltages are so arranged that the ions produced in this space are not counted by the observer. This device is not always used, nor when used is it fully effective. A somewhat experienced man might expect to be able to allow for the radioactivity of the wall, through knowledge acquired by varying the density of the air and studying the ionization as function of density; this has been done, but we shall see that the results have added new mysteries to the problem.

The walls, then, must be expected to cause a permanent ionization in the air-chamber, constant and independent of the outside world. Moreover electrical leaks must occur, whereby the charged electrode gradually loses its charge by conduction through the not-quite-perfect insulators on which it is mounted. How can we hope to distinguish the joint effect of these from that of the cosmic rays? or, to put the same question more properly, how can we hope to find what part of the ionization, if any, is not due to either of these?

This is indeed a serious question, inasmuch as it is a problem not of distinguishing a known from an unknown, but two unknowns from one another. Of course, if one makes the gratuitous assumption that the "cosmic rays" cannot penetrate x feet of rock, then one may take the apparatus into a cave or hollow under more than x feet of rock, and measure the rate at which the charged electrode is discharged; and then, after deducting the allowance to be made for the rays from radioactive substances in the rock which penetrate into the chamber, one can say that the residue (*Restgang*, the Germans call it) is the joint effect of the causes aforesaid, and must thenceforth be subtracted from every reading made with the same apparatus. Kolhörster did this on a grand scale, going many hundreds of metres down into one of the potash-mines of Stassfurt, where there is a great hollow excavated in rocksalt; the thickness of the overlying rock was great, but as a drawback there was a powerful radioactivity of the potassium in the rock-salt, causing hard gamma-rays of which a perceptible fraction invaded the chamber. In Java the physicist Clay took his machinery into a tunnel, where he was covered by rock eighty-four metres thick and far less rich than rocksalt in radioactive atoms; in the Sierra Nevada Millikan descended 185 metres into granite.

More instructive is the customary procedure of Millikan and of Regener. We shall presently consider curves (Figs. 3, 4, 5) which show the rate of discharge of electroscopes in air-chambers sunk under water, plotted against the depth of submergence. The discharge-rate falls off as the air-chamber is lowered, but the decline grows slower and slower; the curve flattens out and seems to approach, seems even to attain, a certain horizontal line. It is the ordinate of this line which is taken as the *Restgang;* this is subtracted from all the other ordinates, the differences are ascribed to the cosmic rays. Now if one could be absolutely sure that this is the line to which the curve is making asymptotic approach, there would be no uncertainty. But since no measurement is perfectly precise, no one can say absolutely that the curve is not still gently sloping, towards an asymptote distinctly lower than the lowest value measured. Yet it is on some guess as to the answer of this unanswerable question that there rest, not (fortunately) the proof that there are such things as cosmic rays, but the estimates of their amount and of their greatest penetrating power. I revert to this question later; at present I merely mention it, in order to show how difficult these estimates may be.

A very important point now demands to be noticed. In taking these elaborate precautions to exclude the rays from radioactive substances beyond the walls of the air-chamber, is one not also keeping out some of the cosmic rays themselves? To deny this would be all but impossible; it would amount to assuming that *all* ionizing rays, apart from those which we recognize as proceeding from known radioactive atoms, are so much more penetrating than these that they pass absolutely undiminished through centimetres of lead and metres of water—an assumption which has only to be stated, to show itself improbable. True, *if* the unknown rays come altogether from above, and *if* the radioactivity of the air can be allowed for, one may omit to carry the screen of lead over the top of the air-chamber—or, one may float the apparatus on or just below the surface of the lake—or, one may rely exclusively on observations made in the upper air. But it is likely that part of the unknown radiation comes very obliquely downward through the atmosphere, and therefore is liable to be obstructed by lateral walls of lead which cannot be removed without admitting to the apparatus other rays which come very obliquely upward from the earth. So, every sort of screening by water or by metal probably keeps out some of the rays which are wanted, along with those which are unwanted. Only the observations in the uppermost air may give the full effect of the mysterious radiation, and perhaps not even they, for the walls of the chamber cannot be reduced to infinite thinness.

How great is the ionization left over to be ascribed to cosmic rays, after *Restgang* and radioactivity are allowed for? The absolute value is still a subject of controversy; fortunately it is a minor matter, knowledge of which is not required for solving the major problems; thus, to determine the penetrating power of the rays one needs only *relative* values at various depths of water. Concerning it I will merely state, that *in the atmosphere near sea-level, the number of ions of either sign appearing per cubic centimetre per second, and attributed to cosmic rays, averages about one and one-half.* Say we have an ionization-chamber of volume U (in cubic centimetres), at sea-level, filled with air at atmospheric pressure. Its charged electrodes will be discharged at a certain rate; part of this rate of discharge can be explained in the ways aforesaid; the rest, the inexplicable residue, will amount to about $1.5\,Ue$. This residue is the evidence for the cosmic rays. The electron-charge e is so small, that even when the volume of the chamber is thousands of cubic centimetres the product ($1.5\,Ue$) is no great quantity of electricity. High up, the effect is much greater; according to Piccard it is so great at sixteen kilometres over sea-level, that if the air up there were as dense as it is by the sea, there would be 200 ions of either sign appearing per cc. per second.

Stating the effect in ions per cc. per second has one disadvantage: it suggests that in each cubic centimetre of an ionization-chamber the ions are created at random,—one in the middle, perhaps, at a certain instant, the next somewhere else two-thirds of a second later, the next yet somewhere else two-thirds of a second later yet, and so on. What happens, however, is this: at practically a single instant hundreds, or even thousands, of ion-pairs are created along a single straight line traversing the chamber; this is followed by many seconds, or even minutes, during which nothing happens; then there is another such event, a train of ions suddenly appearing along a straight line, not however the same one; another interval, another train of ions, and so forth. The method of measurement which thus far I have been presupposing conceals these individual events, giving merely the sum of all the ions which come into being over a period of hours (or, at least, of many minutes). Towards the end of the article I shall speak of other, and very striking, methods which reveal them.

Passing now to the experiments, I will take up first the various groups of measurements made, each with a single instrument of the ionization-chamber or electroscope type, at various heights above the ground or various depths below the surfaces of lakes.

Imagine, to begin with, a metal-walled air-filled box, having inside it a pair of strips of gold-foil hanging side by side from a metal knob at the end of an insulating rod, and in one wall a window through which this "gold-leaf electroscope" can be seen. The metal wall is earthed, and the leaves of gold are charged, usually by a metal rod so mounted in a ground-joint in the wall that by turning the joint the rod can be touched to the knob which holds the leaves. The wall is disconnected from earth, the box is set up in whatever place is chosen for the measurement. The divergence of the gold-leaves is measured at the beginning and at the end of a chosen interval of time. By calibrating the device (I will not enter into the details of this process) one may determine how much charge of one sign has been lost from the leaves, how much charge of the opposite sign has been gained by the leaves, in the time allotted. This is the ultimate datum. The process is then repeated in other places.

All of the experiments which I shall now describe are in principle like this imagined one, but greatly improved in technical detail. Gold-leaves would be too frail for an apparatus meant to be moved about; for them are substituted sometimes a pair of quartz fibres brought together at both ends, not at one only (Fig. 1) or else a "string electrometer." Sometimes the observations are made fifteen kilometres up in the air, or two hundred metres down under water. It is not always convenient, sometimes not feasible, for the observer to go with his instrument; it is not always safe to make the initial reading before the apparatus is sent on its way, the final reading after it is brought back. The observations must then be delegated to a mechanism; the divergence of the fibres or the position of the string is recorded by photography, perhaps at intervals of time previously chosen and regulated by machine, perhaps continuously. The pieces of apparatus designed to do these things, and meanwhile to survive immersion or transportation to great heights and low temperatures, are often prodigies of compactness.

Though the work of Millikan and his school is unrivalled for extent, I will begin with Regener's, which is an unique example of a wide range of water-depths covered in a single experiment. The

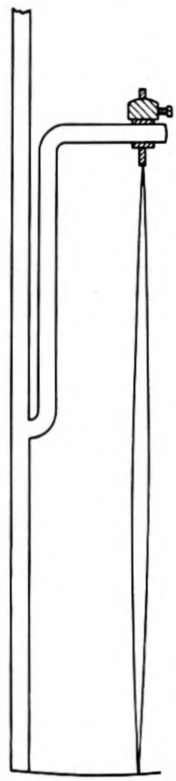


Fig. 1—Sketch of a two-fibre electroscope (after Millikan).

ionization-chamber (walls of steel a centimeter thick, 39 liters (!) in volume filled with carbon dioxide at a pressure of 30 atmospheres) was floated for days at a time at each of several levels in the great Lake of Constance (*Bodensee*) on the northern boundary of Switzerland. Every hour a light flashed for a few seconds, and the image of the fibre of the string-electrometer was impressed upon a film, wound about a slowly and equably revolving drum. One sees in Fig. 2 the successive images of the string, developed after the experiment was completed and the film removed from the drum; it is evident how they lie closer together as the chamber is lowered deeper and deeper beneath the surface, the ionization therefore becoming feebler and feebler. The water, as we shall see, is serving not merely as a screen for keeping out the rays of radioactive matter in the earth but as an instrument for studying the penetrative power of the unknown rays themselves. Records, as I just said, were made an hour apart, and ordinarily were continued over two days at each depth.

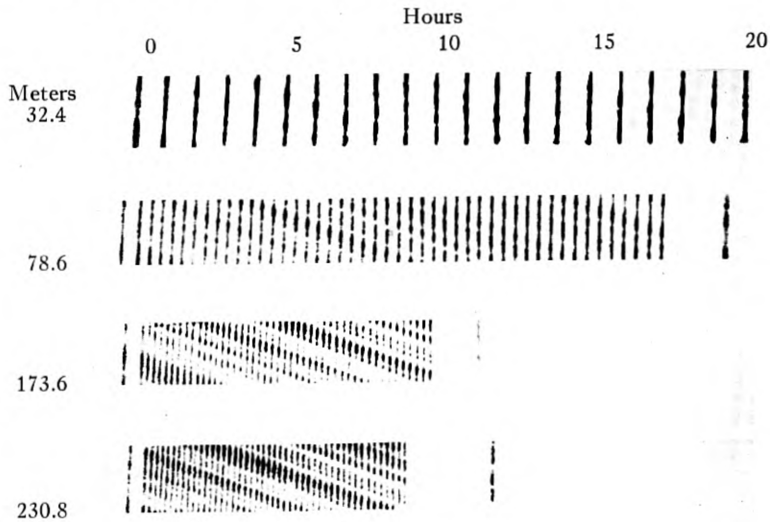


Fig. 2—Photographic registration of the position of the string of a string electrometer, showing gradual shift due to discharge by cosmic rays (Regener).

Measurements were made at seven depths: in metres, these amounted to 32.4, 78.6, 105.2, 153.5, 173.6, 186.3, and 230.8. They were plotted as function of depth; the resulting curve, *shifted downward by 0.78* (we need not consider the exact meaning of the units along the axis of ordinates) appears as J in Fig. 3. It is concave-upward, of a shape which suggests that there are both a *Restgang*, and ionizing-rays diminishing in strength as they descend through the water from above. But it is not sufficiently prolonged to permit the eye to decide with confidence the value of the *Restgang;* the exact effect of the ionizing rays is therefore not to be judged by inspection. How did Regener proceed? He postulated that the readings y at the three greatest depths (depth measured downward from the surface being denoted by x) are sums of a constant and an *exponential* term:

$$y = A + Be^{-\mu x}; \; x = 173.6, 186.3, 230.8,$$

and found what values must be given to A, B, and μ in order to get the best possible fit of such an expression to the data. He then noticed that the same expression would equally well fit the data for the next two of the depths, 153.5 and 105.2; and this convinced him that it is physically sound.

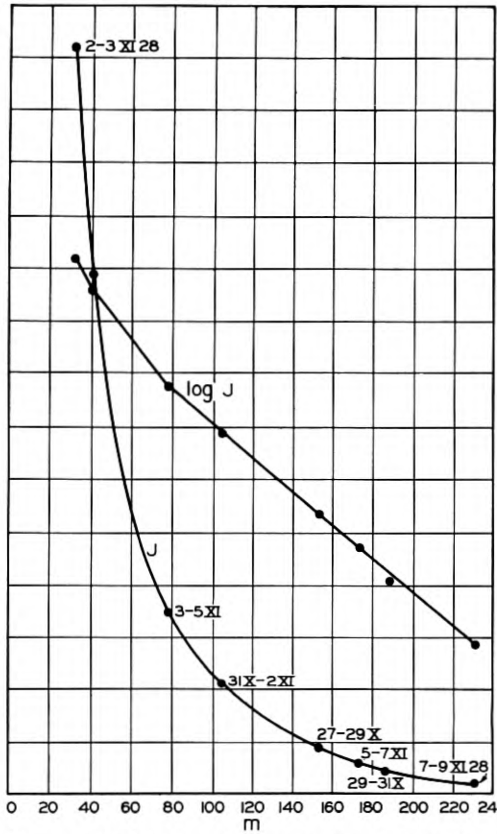


Fig. 3—Ionization ascribed to cosmic rays as function of depth beneath surface of Lake Constance (the straighter curve is a semi-logarithmic, the other a direct plot) (Regener).

The ionization has therefore been separated by assumption into a *Restgang* A and an effect of ionizing rays diminishing exponentially with increase of depth. The value of A is 0.78; it is the downward shift mentioned just above, so that the curve J in Fig. 3 shows the part of the ionization ascribed to cosmic rays. The other curve shows the logarithms of the ordinates of the curve J; the latter being exponential, the former is straight; it is the close fit of this line to the circles along its path (the logarithms of the data) that for the eye is the most striking evidence for Regener's procedure.

Now, of the three known kinds of rays from radioactive substances, one only is characterized by an exponential decline of its strength with increase of thickness of matter traversed: to wit, the gamma-rays. Regener was therefore assuming, as had many others before him, that *cosmic rays which have penetrated many metres of water behave like a beam of monochromatic gamma-radiation, corpuscles of light of a single frequency and energy, coming vertically from above.* The value of μ—.00018 in this case [3]—is many times smaller than that of any known variety of gamma-rays; this is expressed by saying that the cosmic rays are considerably more penetrating or "harder" than the hardest of gamma-rays.

While the expression $(A + Be^{-\mu x})$ fits tolerably the five experimental points at the five greatest depths, it misses the two others, giving for 32.4 and 78.6 metres lower values of ionization than those observed. One might add a second exponential term with a different, much larger value of μ; thus in effect assuming that the cosmic rays behave like a mixture of gamma-rays of two different wave-lengths, corpuscle-energies and hardnesses. This in fact is the custom, though Regener did not do it in the paper whence I am quoting, being interested mainly in the far end of the curve—in the range of depths at which, as is commonly said, *all but the hardest component of the cosmic rays have already been filtered out.* For, if a function of x is the sum of a number of terms of the form $B_i e^{-\mu_i x}$, with differing values of the constants μ_i: then as x is increased the various terms fall off, but in such a way that eventually the one with the lowest μ predominates more and more over all the rest, however small may be its coefficient B. This is "filtering," a process of great use in the science and practice of radioactivity; many of the radioactive elements, or of the mixtures thereof which are common, send out gamma-rays of various degrees of penetrating-power, among which the hardest may be almost isolated from the rest (at the price, of course, of a great cutting-down of its own intensity) by a thick screen of absorbing matter. Even if the screen is not thick enough to make the hardest component predominant, it may alter greatly the relative proportions of harder components and softer. The rays responsible for the 1.5 ions at sea-level, if of the nature of gamma-rays, must have undergone a good deal of this alteration and "hardening" in passing through the atmosphere; not nearly so much, however, as those deep down in Lake Constance, the whole of the atmosphere being no more powerful an absorber than a layer of ten metres of water.

We turn now to the great series of experiments made by Millikan

[3] Expressed in cm.$^{-1}$; in (metres)$^{-1}$, Millikan's customary unit, it would be .018.

and his school. Fig. 4 embodies the data of three, in each of which an air-chamber was lowered into a high-lying snow-fed [4] lake (Millikan's reasons for choosing such I have already stated). Crosses stand for

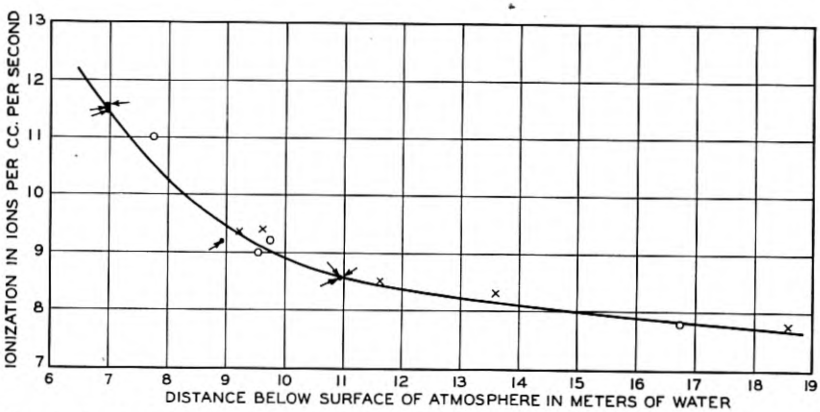


Fig. 4—Ionization ascribed to cosmic rays as function of depth beneath top of atmosphere (early curve of Millikan and Cameron).

data obtained in Arrowhead Lake, California (altitude 1550 metres); circles correspond to Muir Lake, California (3580 metres); the dots to which arrows point are the reward of a long hard journey, for they show data from a lake—Miguilla—no less than 4570 metres above the sea, in the Bolivian Andes. All of the measurements here plotted were taken at depths of a metre or more below the water surface, but none so great as Regener's—five metres was the maximum depth at Miguilla, fifteen at Arrowhead and twenty at Muir.[5]

The points of the three sets lie close to a single smooth curve. Each of the three, however, has been shifted horizontally, to make allowance for the different depths of air overlying the three lakes: it has been assumed that the air over Miguilla is the equivalent of 5.95 metres of water, over Muir to 6.75 and over Arrowhead to 8.6 metres. To figure out the equivalents it is necessary to know the distribution-in-height of the air (Millikan got it from the Smithsonian Institution

[4] Regener of course admits that Lake Constance is not of this character, but by later experiments (briefly mentioned in his note in the *Physikalische Zeitschrift*) he found that radioactivity of its water was not distorting his results.

[5] Data were obtained at lesser depths than a metre, but the values of ionization are higher than would be inferred by prolonging the curve which is valid for greater depths. This recalls a feature of Regener's data, but here the excess is ascribed by Millikan not to soft cosmic rays, but to the radio-activity of the overlying air and the nearby mountains. Of the points plotted in Fig. 4, the "two which fall farthest from the curve correspond to single readings, and hence should be given little weight in comparison with points which represent the means of three or four readings" (Millikan & Cameron).

tables) and the ratio of the absorbing-power of air of given density and the absorbing-power of water for the cosmic rays. This last not being known *a priori*, it is the custom to take the inverse ratio of the densities of the media (page 152)—another instance of assuming that cosmic rays are of the nature of gamma-rays. The assumption is strengthened by the success of the thus-made allowance in bringing all the points upon a single curve.

The data of Fig. 4 were obtained in 1925 and 1926; those of the following Fig. 5, in 1927. New readings were made in the waters of

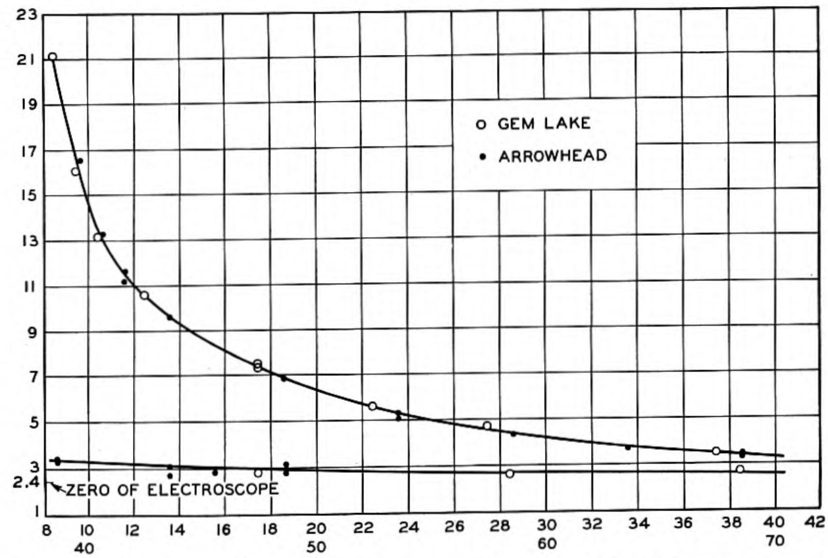


Fig. 5—Ionization ascribed to cosmic rays as function of depth beneath top of atmosphere (later curve of Millikan and Cameron).

Arrowhead Lake and in another not previously used, Gem Lake, 2,760 metres up in the high Sierras; into this last the ionization-chamber was lowered to a depth as great as sixty metres. The data for the two localities were altered to make allowance in the foregoing way for the different heights of air overlying the two lakes, and the values so obtained were plotted against "depth in metres-of-water beneath top of atmosphere." Evidently the data fit better to the curve than in the prior work: this Millikan and Cameron ascribe to improvements in the new ionization-chamber or "electroscope" (outwardly like that in Fig. 6) into which air was compressed to a pressure of *eight* atmospheres instead of *one*, as previously. Incidentally, in Fig. 5 the *Restgang* appears under the title "zero of the electroscope."

The observations had thus been extended over the range, of metres-of-water beneath the "top of the atmosphere," from 8.45 to 69. During the three following years Millikan and Cameron made yet another and more-closely-spaced aggregate of readings, over the same range with slight extensions at both ends, in the same two lakes, with

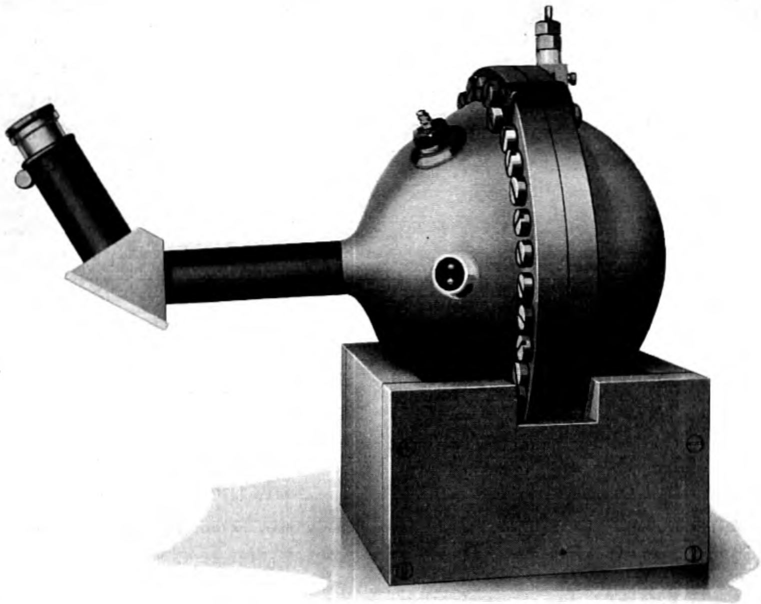


Fig. 6—Electroscope used in later under-water measurements of cosmic rays by Millikan and Cameron.

an electroscope (Fig. 6) into the 1622 cubic centimetres of whose volume air was forced to a pressure of 30 atmospheres. The reader can find the new curve in the issue of the *Physical Review* for February 1931; it is the one from which the next-cited deductions are drawn; but as it does not differ markedly from its two predecessors, and is much more trying to the eyes than these, I do not reproduce it here. (In studying it, or the curve here pictured as Fig. 5, one must remember that the ordinate is proportional to the number of ions appearing per second per cubic centimetre, not of the atmosphere at the level in question, but of air of a standard density.)

Is the whole of the curve, abstraction being made of the *Restgang*, an exponential curve such as would be found if the cosmic rays were a beam of gamma-rays of a single frequency descending vertically from above? The answer presumably is *no*. Millikan and Cameron, however, asked a slightly different question. Imagine that the cosmic rays, instead of coming altogether vertically from above, are gamma-rays traversing interstellar space uniformly in all directions, incident therefore upon the atmosphere at all angles between zero and 90°—much the most probable assumption if they do proceed from interstellar space, as it is hard to think of any possible action of the earth which might constrain them all to move straight toward its centre. The ionization at any depth x beneath the top of the atmosphere (continue to think of x as measured in metres of water) would then be due to beams coming from all directions above the horizon. These beams, having traversed different thicknesses of air, greater the more they are inclined to the vertical, would have been absorbed to different degrees; their intensities, originally (by hypothesis) the same, would be reduced in different proportions. It can readily be shown that the total intensity of all the beams should then vary, not as $e^{-\mu x}$, but as the following function of x:

$$y = C \int_1^\infty z^{-2} e^{-\mu x z} dz, \tag{2}$$

μ standing as before for the absorption-coefficient of any cosmic-ray beam coming from any one direction.[6] Now the curves of Figs. 4 and 5, and the new curve of 1931, do not conform to this relation either. It is consequently not possible to affirm that the cosmic rays behave like gamma-rays of a single wave-length.

Does the final extremity of the Millikan curve conform with formula (2), so that one may assume that after a certain and feasible amount of filtering, the radiation is almost altogether reduced to gamma-rays of a single wave-length? To this Millikan and Cameron respond, that of the curve of Fig. 5 the portion extending from abscissa 30 to abscissa 60 *does* conform, the value required for μ being .0005. Of the curve of 1931 they say, that from 40 to 80 metres-of-water it conforms, if for the value .00028 be chosen.

Recalling Regener's value of μ (.00018) for depths still greater, and joining it with these, one is led to wonder whether at 40 and even at 80 metres the filtering is sensibly incomplete, and whether even at 200 metres the value ascribed to μ may not be the ultimate "hardness" of the rays. Further, there now arises the great question: can the curve as a whole be regarded as the sum of a limited number of terms of the

[6] See R. Hellmann, *Phys. ZS.* **30**, 357–360 (1929); E. Gold, *Proc. Roy. Soc.* **A 82**, 43–70 (1909). Regener used formula (1) instead of formula (2) for interpreting his data.

type appearing in equation (2), with definite and distinctive values of μ? For if so, we may affirm that formally the cosmic rays behave like a mixture of gamma-rays having definite discrete frequencies, a sort of line-spectrum.

This is an extremely delicate question. What is needed is a method like that of Fourier analysis, whereby a given curve may be resolved into a sum, not of sines and cosines as result from the Fourier process, but of terms of the type of the right-hand member of (2); it seems that such a method is wanting. Millikan and Cameron, and independently their colleague Bowen, "built up . . . our observed curve out of four components—no smaller number would do—and in such a way that the synthetic and the observed curve [that of 1931] fitted exceedingly nicely from one to the other." These four components have the values .0080, .0020, .0010, .0003 of the coefficient μ; and their relative intensities, as inferred to exist at the confines of the atmosphere before the filtering commences, stand in the ratios 141000 : 130 : 80 : 33. It is evident that the portion of the radiation which survives at great depths of water, and for which the values of μ are extremely low, is a very small part of all that is to be found in the upper air. Yet even the highest of the values of μ here cited is much lower than the coefficient of absorption for the hardest known gamma-rays from radioactive substances.[7]

It would be deplorable to leave unmentioned some very romantic experiments in which the electroscope, with or without the observer, ascended by the aid of a balloon to heights of air hitherto unattained. Balloon-ascents were begun by German physicists (Gockel, Bergwitz, Hess, Kolhörster) in the five years before the war. Before, it had been maintained that the ionization of the air is due altogether to radioactive substances in the ground. The earliest balloon-flights impaired this argument, by proving that the amount of the ionization does not diminish rapidly as the observer flies upward; the later ones destroyed it, by proving that above a certain level (variously stated, and depending no doubt on the thickness of the walls of the chamber) the ionization rises with increase of height. I will however speak chiefly of more recent flights, those of Millikan's apparatus and that of Piccard.

The electroscope of Millikan and Bowen was borne aloft above the plains of Texas by a pair of balloons, one of which eventually burst and left the other to serve as a parachute in lowering its burden gently

[7] Though not higher than that of the gamma-rays excited by Bothe and Becker when bombarding beryllium with alpha-particles, as I learn from a letter of Professor Bothe.

back to earth. In two experiments the electroscope climbed to heights of 11.2 and 15.5 kilometres, heights to which it seemed (until last year) that no man could possibly ascend and live. The apparatus (there were four of them) had to make its own records and bring them back; it was a masterpiece of ruggedness and compactness jointly, as one sees in Fig. 7, on the left of which it appears assembled (with a

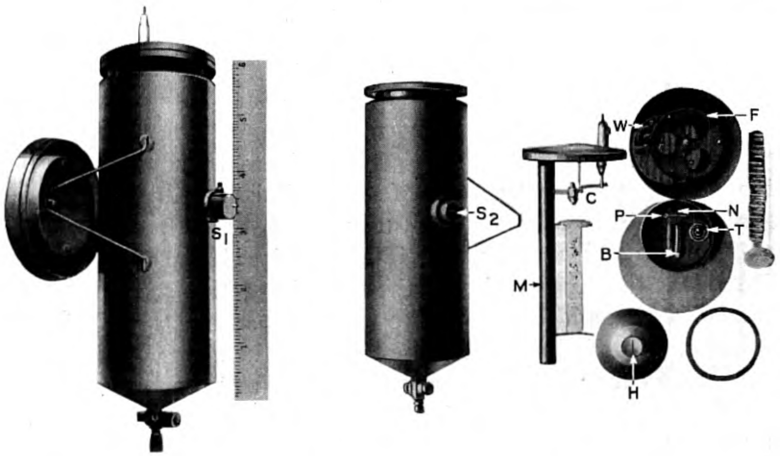


Fig. 7—Electroscope used in high-altitude measurements of cosmic rays by Millikan and Bowen (photographs somewhat retouched).

six-inch rule beside it to show its scale), on the right resolved into *disjecta membra*. The symbol B marks a manometer in the form of a U-tube with one arm closed, containing liquid; the symbol T, a thermometer in the form of a coil of metal with a pointer at its end. The light of the sky imprinted images of the meniscus of the barometer-liquid and the pointer of the thermometer-coil upon a moving film, kept in motion by the watch of which the mechanism is marked by W. Inside the cylinder M (so at least I read the text) were the two fibres of the electroscope already shown as Fig. 1, and they also were continually photographed by the light of the sky upon another film which the same watch kept in motion.

When one of these devices had ascended to 15.5 kilometres (nearly ten miles) and returned to earth eighty miles away from its starting-point, and had been picked up (by some casual wanderer, one infers) and returned to its authors, the films bearing the records of temperature and fibre-divergence looked as they do in Fig. 8. To determine the extent of the ionization at great heights without being deceived by the influence of temperature on the fibres, Millikan and Bowen compared the divergence of these when the balloon had reached a certain height (it was five kilometres) on the upward way, with their divergence as the apparatus passed the same level in descending. During the time while it was above this level, the fibres of

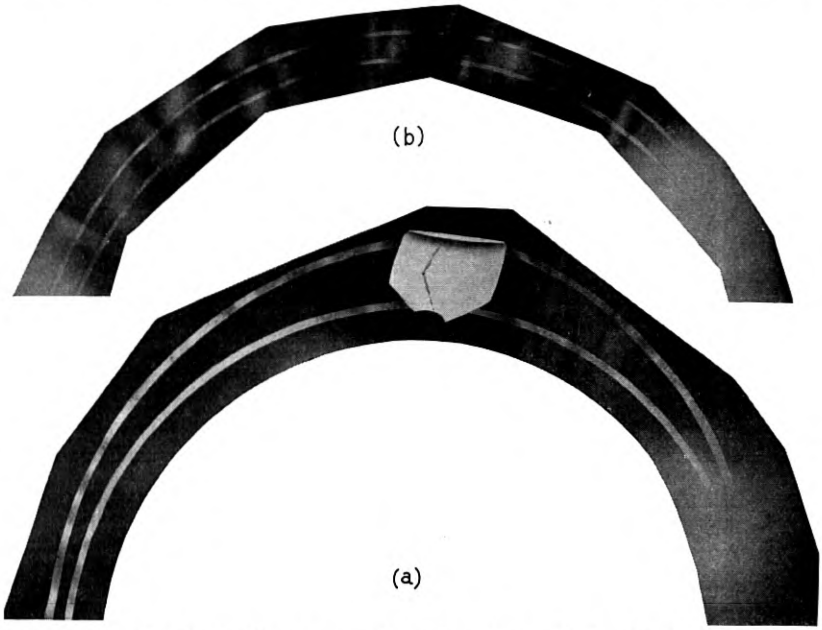


Fig. 8—Records obtained by automatic registration at high altitudes by Millikan and Bowen (somewhat retouched).

the electroscope lost three times the charge that would have leaked away, if the instrument had been left on the ground. This again is proof that the cosmic rays are stronger in the upper air than in the lower.

As I remarked above, until last year it had been deemed beyond the powers of living man to soar to heights so great as fifteen kilometres. This opinion was confuted by the superbly audacious flight of the Belgian physicist Piccard, which everyone will remember who reads the papers, though other things than cosmic rays were stressed in their accounts. Piccard seemingly has not yet published a full nor any extensive story of his measurements on cosmic rays, not at any rate in the physical journals; but in the bulletin of the French Physical Society I find an item in which he states as *provisional* result—awaiting further calibration of his instruments—that the ionization at a height of sixteen kilometres, in air at standard pressure, is of the order of 200 ions per cc. per second! This is considerably greater than the modest value of 1.5 prevailing at the level of the ground.

I must make at least a passing allusion to the *fluctuations* of the intensity of the cosmic rays—fluctuations which are smoothed over by some of the experiments and missed by others, for they are too small to be certainly detected by measurements of low precision, and many are too rapid to be noticed by measurements on the amount of discharge of an electroscope over a period so long as one or two hours.

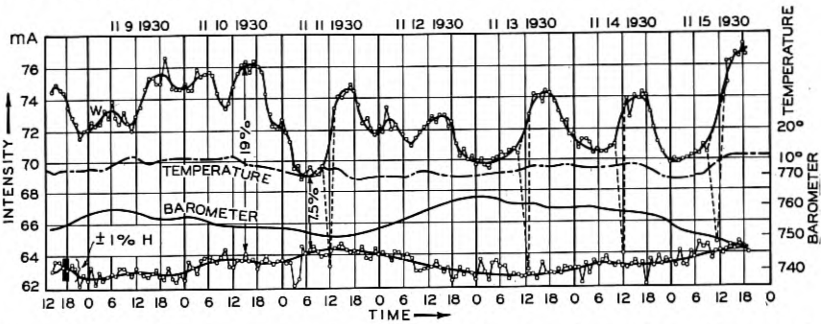


Fig. 9—Fluctuations of cosmic-ray intensity and attendant fluctuations of barometric pressure and temperature (Hoffmann).

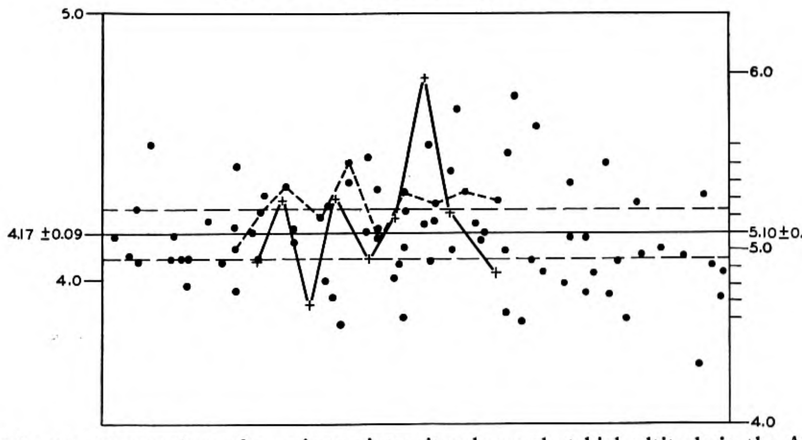


Fig. 10—Fluctuations of cosmic-ray intensity observed at high altitude in the Alps (von Salis).

Figs. 9 and 10 display examples of these: to appreciate the ratio of their average amount to the mean value of the ionization, one must take note of the scale of ordinates on the left. The former is especially instructive, for both the uppermost and the lowermost of the curves display the ionization of the air near the ground (in a basement at

Halle) measured at short intervals over the length of a week; the difference is, that the lowest shows the effect inside armor of lead ten cm. thick, while in getting the data for the topmost the plates of lead were removed from above the apparatus, though left in place at the sides and below. The lower curve thus represents the action of hard cosmic rays: one sees that it rises as the barometric pressure falls, falls as the latter rises—the heavier the blanket of air between Halle and the sky, the more these hard rays are reduced on their way to the ground. The upper shows the joint action of the hard cosmic rays, of others which are softer, and of soft rays from radioactive substances in the air: one sees how violently it fluctuates, supposedly (in Hoffmann's view) because the radioactive atoms wandering in the air change greatly in number as the weather changes. Figure 10 shows values of ionization obtained high up (3500 metres) in the Alps, on various August and September days of 1927, plotted against sidereal time; the zigzag lines connect observations made on two particular days. Corlin is said to have inferred, from a statistical study of data obtained in northern Sweden, that the ionization ascribed to cosmic rays decreases gradually and slowly before a magnetic storm, leaps suddenly to a high value at the onset of the storm, then decreases again.[8]

There is a related question which has been much debated: does the intensity of the cosmic rays vary as different celestial objects pass overhead, the sun for instance, or the Milky Way? As the years go on, the answer to this question becomes steadily more and more strongly in the negative. The opposite opinion has been held largely by German physicists; but a year ago one of them (Hess) reduced the proportion of the cosmic rays which he considers dependent on the sun, to half of one per cent of the total amount.

Certain perplexing data must be mentioned before we go on to the work which is done with other instruments than the ionization-chamber, for they impeach the reliability of this device. From many experiments of the pupils of Swann, it appears definitely established that if the quantity of air in an ionization-chamber is increased, the number of ions appearing in that chamber in unit time increases in a lesser ratio, and in fact approaches (and in some practical cases, even attains) a limiting value (Fig. 11). At first this seems natural enough: the limiting value should be attained when the air is so dense, that the ionizing rays are entirely absorbed before they have completely crossed the chamber. *But*, appreciable fractions of these ionizing rays are able to penetrate many metres of water, many centimetres of lead: it is pretty nearly a formal contradiction-in-terms, to assert that they

[8] See W. M. H. Schulze, *Nature* **128**, 837–838 (14 Nov., 1931).

can be fully absorbed in air of the density and thickness involved in these experiments. It seems necessary to suppose that as the pressure is increased, an ever-rising fraction of the total number of generated ions fails to make its way to the electrodes, positive ions and negative

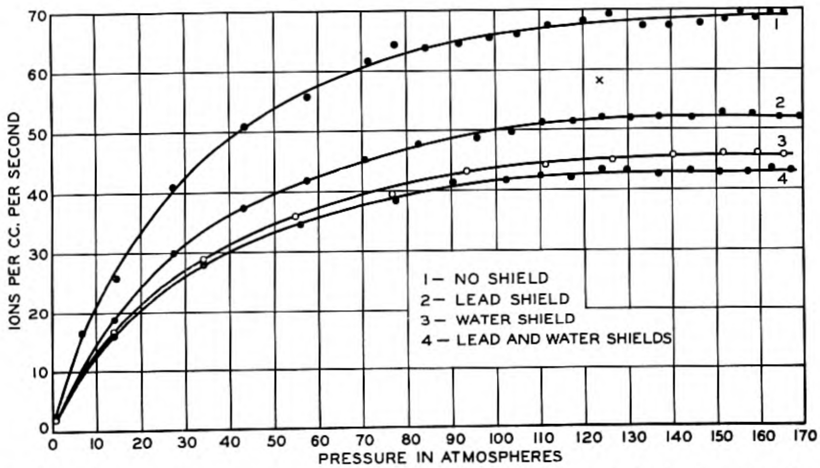


Fig. 11—Curves of ionization versus quantity of contained air in ionization-chamber (Broxon).

recombining with one another.[9] If so, there is reason to fear that the readings of a high-pressure ionization-chamber are affected by its temperature—a possible cause of spurious fluctuations.

I turn now to the other ways of experimentation: there are two of them, both furnishing data more striking and spectacular than those of electroscopes. They exhibit the immediate cause of the mysterious ionization, and open the question of its ultimate cause. We must now take up this question: is the *ultimate* cause to be compared with electromagnetic waves or gamma-rays descending from above, as in this article I have thus far let be assumed? or, is it a rain of fast-flying material particles such as electrons or protons? or a mixture of the two classes of radiation? or has it properties unlike those of any radiation so far known?

(This question is often put in the words: are the cosmic rays wavelike or corpuscular? I think that this phrasing is very much to be condemned, gamma-rays as well as material particles being at one and the same time *both* wavelike *and* corpuscular. However the reader should know that in current usage "corpuscular" signifies "composed of particles of electricity or matter, instead of photons.")

[9] This theory, almost simultaneously published from Millikan's school and from A. H. Compton's, is at present under test.

The question is much more difficult to answer than it seems: for *even if the cosmic rays are of the nature of gamma-rays, the effect which they produce in ionization-chambers (and in the kinds of apparatus yet to be described) is almost entirely due to fast electrons.*

Consider a beam of X-rays (of the same nature as gamma-rays but softer) projected into a gas. It is known by experiment that they free electrons from atoms in two distinct ways: some of the corpuscles or photons are entirely absorbed in atoms, and tightly-bound electrons are ejected (photoelectric effect); some on the other hand make elastic impacts against loosely-bound electrons, communicating a part of their energy to these and retaining a part (Compton effect). As the frequency of the X-rays is increased, the former effect becomes less usual; entering into the range of gamma-ray frequencies, we find it fading out; it is reasonable to suppose that if there are corpuscles of light of frequency even higher than that of the gamma-rays, they ionize only by Compton-effect or "Compton scattering." It remains of course conceivable that such corpuscles may ionize in ways unknown at lower frequencies, as for instance by shattering the nuclei of atoms into fragments some of which are fast-flying electrons or protons.

Now though these impacts are properly called "elastic," their laws are peculiar, owing to the differences between corpuscles of light and corpuscles of matter.[10] The electron acquires some, usually a considerable part, of the energy of the photon; this latter (or as some would prefer to say, a new photon) goes off more or less obliquely, with lessened energy and lessened frequency. Given the wave-length of the light, the laws of the impact determine the energy acquired by the electron, the energy retained by the photon, and the direction of departure of the photon, as functions of the direction in which the electron is projected. If we know the distribution-in-direction of the projected electrons, we can determine their average energy. For electrons ejected by X-rays this distribution is known. But if the cosmic rays are (or include) gamma-rays whereof the absorption-coefficient has such values as those given (*e.g.*) by Millikan and by Regener for μ, they are much harder than any otherwise-known gamma-rays or X-rays, hence presumably of much lesser wave-length. Then, for the distribution-in-direction of the electrons which they eject, we must resort to an untested theory. This theory indicates that a beam of short-wave gamma-rays proceeding through matter should provide for itself an escort of electrons, most of them moving in directions inclined at very small angles to the beam itself, and many

[10] See for instance my *Introduction to Contemporary Physics*, pp. 145–163, or the sixth article of this series.

of them starting out with values of kinetic energy which are considerable fractions of the energy of the original photon.[11]

It is this escort of fast electrons which should be the immediate cause of the ionization due to the short-wave gamma-rays. One easily sees how this result is destined to complicate the question of the nature of the cosmic rays. If it can be proved that the ionization attributed to these is caused by fast electrons, this may mean that the electrons are coming into the lower atmosphere from above, in which case most physicists would say "the cosmic rays are fast electrons." On the other hand it may mean that the electrons are produced in the lower atmosphere by photons coming from above, in which case most physicists would say "the cosmic rays are of the nature of gamma-rays." Yet in both cases the immediate agent would be the same; the term "cosmic ray" would be applied in the one case to it, and in the other to the ultimate cause of the immediate agent. This fine distinction, partly physical and partly verbal, is the source of a lot of confusion.

Consider now the evidence about the nature of this immediate agent. About three years ago the physicist Skobelzyn was engaged in studying the Compton scattering of gamma-rays from a certain radioactive substance, using the "expansion" or "cloud-chamber" method. By this method (it has often been described elsewhere,[12] I will therefore omit the details) the ions formed in a glass-walled container of gas are in effect rendered visible, each becoming the centre of a droplet of water, which appears on a photograph taken when the gas is illuminated. If during the experiment the chamber is traversed by a fast electron, its path is marked out by an unmistakable train of droplets. If a strong magnetic field is applied to the gas meanwhile, the trails of the fast electrons are visibly curved, and their speeds may be computed from the curvatures and the field strengths. Skobelzyn applied a field of 15000 gauss, in which the paths of the electrons liberated from atoms by the gamma-rays were curled up into beautiful spirals. But on examining the six hundred and thirteen photographs which he took, he found twenty-five trains of droplets resembling the curled ones due to the electrons of known cause, but not perceptibly curved at all! and two others of which the curvatures were perceptible, but so slight

[11] See for instance the article of H. Kulenkampff (*Phys. ZS.* **30**, 561–567; 1929) where he plots distribution-curves deduced by the quantum-mechanical theory of Klein and Nishina for electrons ejected by light of wave-lengths $2.4 \cdot 10^{-13}$ and $24 \cdot 10^{-13}$. He also considers the influence of the photons in continuing to eject new electrons after their first, second and later impacts; it appears that one photon is likely to start off several electrons at various points of the beam, instead of disappearing (so far as ionizing-power is concerned) after its first impact.

[12] As for instance in my *Introduction*, pp. 45–46, or the first article of this series.

that out of them he estimated the enormous values of 7 and 15 millions of equivalent volts for the energy of the two particles. As for those of which the tracks were not sensibly curved by the field, the energy of some at least (assuming them to be electrons!) must have exceeded 15.10^6 equivalent volts. (Of the rest, the paths were so unfavorably placed that absence of sensible curvature might have been compatible with energy-values as low as, but no lower than, 3.10^6.)

This method has also been adopted by Millikan's collaborator, C. D. Anderson, who has a magnetic field of great extent, pervading a large expansion-chamber so oriented that particles coming along or near the vertical (we shall see evidence that near sea-level, the particles do favor that direction) are subjected over a long distance to its deflecting power, and electrons with energy values amounting to scores of millions of equivalent volts are sensibly deflectible. On some of his plates there are trails curved in the right sense for electrons coming downward from above, with energy amounting to 70 millions. On some there are trails curved in the opposite sense; if they were made by electrons, these must have been travelling from the earth upwards; if they were made by descending particles, as seems more plausible, these must have been positively-charged. On one there appear three paths, two apparently springing from a common origin near the wall of the chamber; one is curved in the proper sense for an electron, one in the opposite sense—if it is the track of a proton, this must have had energy of 120 millions—the third is sensibly straight.[13] Skobelzyn too had got plates on which two or three tracks appeared, coming probably from a common point of departure.

It appears from these pictures that the immediate agents of the ionization ascribed to "cosmic rays" are able to produce long trains of ions closely crowded together (the number of ions per centimetre is probably of the order of one hundred) [14] which is so far as our experience runs, a distinctive feature of electrified material particles such as electrons and protons; that some are deflected in practicable magnetic fields; that from the deflections it probably follows that the charge is sometimes negative and sometimes positive (the uncertainty being due to the fact that from the curvature one cannot tell the sign of the charge unless one knows in which sense the particle is going along the path); that the ones which are known to be charged have enormous

[13] I am much indebted to Dr. Millikan for showing and interpreting these plates to me. They have been shown in scientific meetings and mentioned in the press; the publication of the pictures and the work in the scientific journals will be eagerly awaited.

[14] The values most highly esteemed are obtained not by counting droplets, but by using the third method to determine the number of ionizing particles, the first to determine the total ionization; Kolhörster and Tuwim give $135 \pm 10\%$.

energies, the others (those of which the paths are uncurved) either have yet more enormous energies or else are neutral. These facts make it impossible to suppose that photons are the immediate agents. Before continuing with the deductions we will consider the third method.

The third method of studying cosmic rays involves the use of a "Geiger counter" or some modification thereof—a metal-walled gas-filled tube with a needle or (more commonly) a wire mounted inside it and insulated from the wall, which is connected to one pole of a battery, the needle or wire to the other. If the tube is properly designed and treated, the E.M.F. of the battery properly chosen (this seems to require a lot of experience) the passage of an ionizing particle across the gas between needle and wall is likely to evoke a sudden and violent and very short-lived current-flow. If a telephone-receiver is in the circuit, there is an audible click; if a galvanometer or electrometer is used with an optical device for recording its deflections upon a moving film, there will be a photographic trace of the discharge.

In experiments on cosmic rays with this device, it is the custom to employ a pair of counting-tubes (or even three) and to accept as valid data only the discharges which occur in either simultaneously with

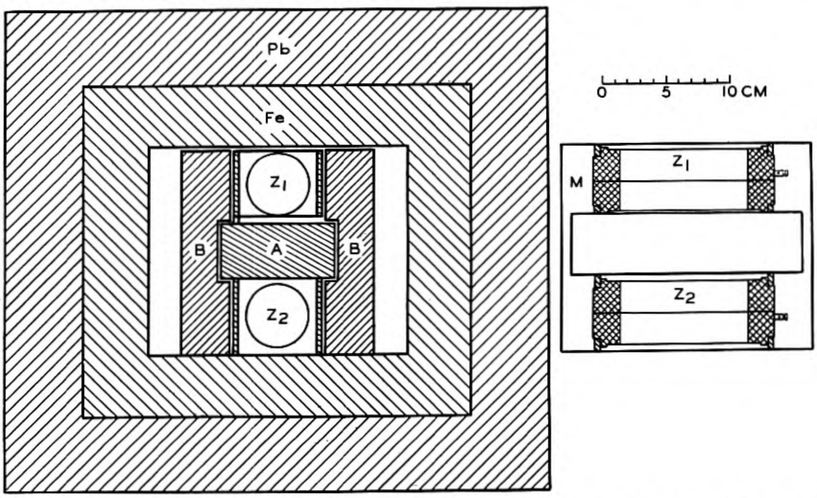


Fig. 12—Apparatus of Bothe and Kolhörster for counting ionizing particles and estimating the absorption thereof in metals.

discharges in the other. I illustrate the method by a sketch from the work of Bothe and Kolhörster (Fig. 12) in which on the right one sees the two counters in longitudinal section, on the left one sees them in cross-section as they lie within their heavy armoring of metal.

"Coincidences"—which is to say, pairs of discharges occurring so close in time to one another that they seem to be simultaneous, that is, within an interval which with good apparatus may be a hundredth or even a thousandth of a second—are much more frequent than they would be, if they were due only to chance. Some of course *are* due to chance: these are the "spurious" or "casual" or "accidental" coincidences, of which the number must be estimated and subtracted from the number observed.[15] After this is done, one is pretty safe in assuming that the remainder is a measure of the number of ionizing particles which have darted through *both* of the counters. By this drastic procedure, of course, the observer limits himself to the study of those corpuscles which happen to be moving along paths which intersect both of the chambers. There will also be corpuscles of which the paths traverse one of the counters, but not the two; it seems as though the observer were throwing away his opportunity of studying these, by rejecting the non-coincident discharges; but some of the discharges in either tube are probably due to radioactive substances in its walls, or to unknown causes inherent in the tube and not connected with cosmic rays; and by accepting only the coincidences, one guards to a great extent against being misled by these. Moreover, by varying the relative position of the counters—for instance, by putting one of them first above and then beside the other—one may study the distribution-in-direction of the ionizing particles. Such studies have not yet been plentiful; but it has been found that if the counters are placed one above the other, the coincidences are several times more numerous than if they are placed side by side.

We have just been considering the effect of magnetic field on the tracks of the corpuscles ascribed to cosmic rays, observed by the cloud-chamber method: it is suitable now to review what the counting-tubes

[15] Suppose that there are N_1 discharges per unit time in one of a pair of counters, N_2 per unit time in the other; and that if a discharge in either starts within a time T after the starting of a discharge in the other, the records of the two (whatever be the way of recording them) make them appear simultaneous. Then in unit time, the periods during which any discharge occurring in the first counter is not separately recorded add up altogether to an amount $2N_2T$. This is the fraction of time, discharges occurring during which are recorded as coincidences. If the events in the two counters are entirely uncorrelated, then out of N_1 discharges occurring per unit time in the first, $N_1(2N_2T)$ or $2N_1N_2T$ (on the average) will coincide with discharges in the second. If the discharges of each counter are recorded as peaks on tracings of separate films, one may determine N_1 and N_2 from the films, set for T the greatest time-difference which cannot surely be distinguished from zero and subtract $2N_1N_2T$ from the total number of coincidences. Or the number of accidental coincidences, together with N_1 and N_2, may be observed by placing the counters so far apart that no particle can traverse the two of them, and T thence computed. Or both sets of discharges may be recorded on the same tracing by connecting both counters to the same electrometer, in which case T will be the greatest interval for which two peaks merge apparently into one; or they may be so connected that no record whatever appears unless two discharges coincide.

have to say about that question. Of the three physicists who have announced experiments upon the problem, L. M. Mott-Smith has given the most impressive account. There are three tubes, of the type exhibited in Fig. 12, their cylindrical chambers being some four cm. wide. Two are in the same vertical line, 30 cm. apart; the third moves to and fro along a horizontal line which intersects that vertical, 35 cm. below the lower of the two which are stationary. Triple coincidences are counted; they are most frequent when all three of the tubes are in line, as one would expect; when the movable tube is shifted away from the line, the number declines, the curve of coincidences N versus displacement x having "about the shape which is expected from the geometry of the arrangement, assuming rectilinear passage of the ionizing particles."

There is another object in the vertical line beneath the lower of the stationary tubes, which I have not yet mentioned: it is a piece of iron 15 cm. thick, which in some of the experiments is unmagnetized, in some is magnetized in such a sense that charged particles which have descended vertically through the upper counters must be deflected, while remaining in that vertical plane which contains the horizontal line along which the movable tube is being shifted. If the particles are charged and of a single speed, the curve of N versus x should be shifted along the axis of x; if they are electrons of energy 10^9 (equivalent volts), it should be shifted by 2.2 cm., and should be easily distinguished from the curve obtained when the iron was demagnetized. Nevertheless, the two curves (or rather the three, for the iron was magnetized first in one and then in the opposite sense, in the hope of producing opposite deflections) are indistinguishable. Mott-Smith considers that if the ionizing particles be electrons, their energy must be not less than 2.10^9—a lower limit which may be halved, if they be protons. These values however are contingent upon an assumption, which I will mention in a moment.

The others who have entered upon this problem are Curtiss and Rossi. Rossi's first apparatus was remarkably like that which I have just described, independently and far apart though he and Mott-Smith were working—the one in Florence, the other in Texas. Later he modified the scheme, having two counters only, and a pair of slabs of iron so placed that when they are magnetized in one sense charged particles passing through the upper counter and thence into either slab should be deflected towards the lower counter, and when they are magnetized in the other sense the charged particles should be deviated away from the lower counter. Again the result was negative; and, says Rossi, "the corpuscular rays are not deflected . . . to such a great extent, as electrons or protons of 10 or 20 times 10^9 equivalent volts would be."

Unfortunately this, like Mott-Smith's, estimate is based upon a tentative answer to a question, one of the most perplexing in physics, though not so familiar as it should be to many who are interested in radiation, ionization of gases, and such phenomena: it is the ancient conundrum, *What is the strength of the magnetic field inside of a magnetized solid?* [16] Mott-Smith and Rossi both assume that the said field strength is equal to the induction B, admitting however that another assumption might be made. If instead one were to put it equal to the "magnetizing field" H, the lower limit inferred for the energy of the corpuscles would sink to a value so small, that the experiments would scarcely even be interesting. Considering this, it is especially important to recall the experiment of Curtiss which preceded the two others. He used two counters only, so foregoing the advantage of the narrow vertically-descending beam to which Mott-Smith confined his study; but between them he placed, not the iron of a magnet, but the space between the pole-pieces thereof; and he *did* get a positive result. The fieldstrength was 7000 gauss, the field extended over 24 cm. of the distance between the counters; vertically-descending electrons traversing the upper counter should have been deflected away from the lower, had their energy been 10^9 or less; actually "making allowance for accidental coincidences, a decrease of the order of 25 per cent has been observed."

We now take up a great but very thorny question: what do counters testify about the penetration and absorption of the cosmic rays in air, in water, and in metals?

Counting-tubes have been taken to great heights of air by Piccard, lowered to great depths of water by Regener. Neither has published (so far as I know) a full account of his data; but Regener implies that the curve of number-of-discharges-per-unit-time (it seems that one tube alone was used) versus depth-of-water agrees in shape with the curve of Fig. 3, especially in respect of its lowest end with the least value of μ. Bothe and Kolhörster took a pair of counters four hundred metres down into a mine, and found—if I read correctly—that there were no coincidences at all, apart from casual ones; a valuable result! The most striking work heretofore published was done in another way: by putting a thick piece of lead (or iron, or gold) either above or between a pair of counters set up in the same vertical line, and ascertaining in what proportion the coincidences are cut down.

The first who tried this out were Bothe and Kolhörster. In the

[16] I have discussed this question (without professing to solve it) in this journal, **6**, pp. 295–310 (1927). Or, see any textbook of magnetism.

particular experiment which I will quote, the apparatus of Fig. 12 was set up under a skylight on the top floor of a building not far above sea-level (the Reichsanstalt in Charlottenburg) *without* the upper part of the metal armor, so that rays coming vertically from above should reach the upper counter after passing through the atmosphere and a little glass. Between the upper and the lower counter, a block of gold 4.1 cm. thick was alternately inserted and removed. Over a period of twelve half-hours with the gold-block *absent* there were 19897 discharges in the upper counter, 12209 in the lower, 986 coincidences; over a period of twelve intercalated half-hours with the gold-block *present* there were 19814 discharges in the upper counter, 10562 in the lower, 743 coincidences. The presence of the gold had thus reduced the coincidences by 24.6 per cent; this the authors qualify by a considerable statistical uncertainty, stating it as (24.6 ± 4.2) per cent.

Rossi did the like experiment, with lead and roofing over the upper counter, equivalent in thickness to 77 cm. of water; in place of the gold he used a slab of lead 9.7 cm. thick; it reduced the number of coincidences by sixteen per cent, which he expresses as $(16 \pm 3)\%$. Only a week or so before I add these words, he reported yet another.[17] There were three counters in a vertical line, 7 cm. of lead and 6 cm. of roofing above them, 25 cm. of lead permanently between them; and an additional thickness of *no less than seventy-six cm.* of lead, which he could interpose between the counting-tubes (half of it between the upper two, half between the two lower). This enormous absorbing layer cut down the number of (triple) coincidences by only some forty per cent (Rossi writes 38.5 ± 5.1).

Now, the coincidences are ascribed to ionizing particles such as those of which the tracks are made visible by the cloud-chamber method: particles which make densely-crowded trains of ion-pairs along their paths. Let us forget for the moment about high-frequency photons, and think only of such particles, descending almost vertically from on high. The numbers 24 and 16 and 38, quoted above, are the percentages of these which are stopped by plates of gold and lead, of the stated thicknesses, in the situations stated. Now the remarkable and perplexing point about these numbers is, that they are about the same as the percentage declines of ionization which are found, by Millikan and Regener, on lowering an electroscope through the equivalent thicknesses of water![18]

[17] I have just received (February 1) a preprint of the report from Dr. Rossi; it is marked *Naturwissenschaften* **20**, 65 (1932).

[18] Bothe and Kolhörster, and Rossi, elect the following way of making the com-

At first glance this appears a simple and welcome result. Further study, however, dispels the pleasant impression: this is not an instance of the sort of observation which sometimes comes aptly to confirm a simple theory and smooth the path for further progress; it is instead a warning of dangerous complications. In introducing the curves of Figs. 3 and 4 and 5, I followed the policy of speaking of them as absorption-curves for rays analogous to gamma-rays descending from above; assuming in effect, that any ordinate of such a curve is a measure of the strength of the gamma-ray beam at the corresponding level of air and water. But if such rays there be, they ionize not directly, but by means of fast-flying electrons (or perchance, other particles) which they expel from atoms. A photon does not achieve the whole, nor even the most, of its ionization at the point where it dies; the effect is spread all along the path of the electron which is heir to a great part of its energy; and this path may be long. Thus, the ionization at a certain level of water or air may be due in part, it may even be chiefly due, to electrons which were unleashed at levels much higher. It may thus be chiefly due to photons which ended their careers far away from the place where it is being measured. It may then be no true measure of the number of photons which collide with electrons at the level in question, and no true measure of the strength of the gamma-ray beam at that level. Indeed it is not likely to be a true measure, unless the range R of the electrons be so short that the strength of the beam does not sensibly diminish as it descends through a distance equal to R. (There is an exception, as I mention on the next page; but in practice it may not help.) Now precisely this condition according to the testimony of Bothe and Kolhörster and Rossi, is *not* fulfilled.

Since the condition apparently is not fulfilled, all sorts of difficulties arise. If the ionizing particles have such ranges and such powers of penetration as the data suggest—not only the data just mentioned, but those of the cloud-chamber method and yet others—many of those which cause the ionization at any particular level of the air have come from points hundreds of metres, or even kilometres, farther up. Any ordinate of such a curve as that of Fig. 4 is a function of the whole of the (unknown) curve representing strength-of-gamma-ray-beam as

parison. Denote by I_0 the number of coincidences per unit time found with nothing between the counters, by I the number found with the block of gold (or lead) between, by d the number of grams per square cm. of gold (or lead). They equate I/I_0 to $exp\ (-\mu d)$, and compute the value of μ; in the first and the third of the experiments just mentioned, it comes out equal to $(3.5 \pm 0.5)\cdot 10^{-3}$ and $(1.6 \pm 0.3)\cdot 10^{-3}$; these are values of the same order of magnitude as those given by Millikan and by Regener for the quantity called μ in the analysis of their curves. In view of the objections to assuming exponential absorption-curves for rays which are or may be composed of fast-flying electrons, I doubt whether it is expedient to use this way of stating the results.

function of height, over an interval of this magnitude. This makes the problem of determining the trend of that curve (if such a curve there be) a formidable one.

Since the problem is so formidable and so important, let us linger over it. It is expedient to treat the very simplest conceivable case, a case which is certainly much simpler than reality, yet quite complicated enough for a first view. Suppose a beam of photons of a single wave-length descending vertically, first through vacuum, then (at $x = 0$) entering a horizontal sheet of some medium (water, say) of uniform density, in which they have an absorption-coefficient μ. Assume that every impact of a photon against an electron results in the total disappearance of the former, in the projection of the latter straight onward (vertically downward). Assume further that the projected electron engenders a constant number of ion-pairs per centimetre of its path, and that the length R of the path is the same for all the electrons.

Then, one readily sees that the ionization in the water, instead of being greatest at the top and diminishing steadily downward, actually increases from the top down to the depth R, and begins to decrease only beyond R. The formula is as follows, x standing for distance measured from the surface of the water downwards:

$$\begin{aligned} I &\propto 1 - e^{-\mu x} \qquad &\text{for} \quad x < R, \\ I &\propto e^{-\mu x}(e^{\mu R} - 1) \qquad &\text{for} \quad x > R. \end{aligned} \tag{3}$$

There is a "zone of transition" of thickness R, beyond which the ionization diminishes exponentially, with the same exponent as we have assumed for the absorption of the gamma-rays themselves. This is the exceptional case mentioned above, in which by diving to depths exceeding R one could arrive at a region where the trend of the ionization is the same as that of the strength of the beam of photons, and the value found for μ would apply to these last.

Unfortunately there is no actual case so simple. If there are cosmic gamma-rays coming from the sky, they evidently come from all directions, not merely vertically; and from the character of Millikan's curves it seems likely that they are of several or many wave-lengths, not one only. Conceivably at great depths of water the hardest of all may be filtered out, and these depths may be superior to the values of R for all the electrons; if so then the lowest values of μ, quoted by Millikan and by Regener, may pertain to actual photons. If there are cosmic gamma-rays of several frequencies descending vertically from above, then the actual ionization-curve should be a sum of terms such as (3), with numerical factors depending on the frequencies and the intensities, and values of μ depending on the frequencies; in principle, one might resolve the curve into these terms; but there is a great difference between "in principle" and "in practice." If instead of water we consider the atmosphere, there is reason to infer that the range R of the ionizing corpuscles discovered in cloud-chamber and counting-tube experiments may be greater than the thickness of the atmosphere itself! In this case, however, there should be a maximum of ionization lower down in the air, than the heights already attained by balloons; such a maximum has not been found. The possibilities are enormously complex; and after mentioning that Steinke has made a number of measurements with sheets of metal of thicknesses apparently inferior to the range of the electrons in the metal, with significant results (he speaks of *Uebergangseffekte* or "transitional effects"), I will pass to the description of an experiment by Rossi meant as a contribution to this problem.[19]

I have already quoted the value which Rossi obtained for the percentage drop in the number of coincidences occurring when a plate of lead is slipped between two counters—16 per 100. When he slipped the same lead plate above the upper counter, the drop was less—12 per 100. Let us imagine that photons coming from above were releasing fast forward-flying electrons from atoms in the plate of lead, these passing in the latter case through both of the counting-tubes, in the former through only one—causing coincidences therefore in the latter case, not in the former. The numerical values then oblige us to suppose that when 100 fast electrons dash from above into the upper counter, 16 of them can be stopped by ten centimetres of lead, but accompanying photons will generate four new ones in the slab, leaving a net decrease of 12 per 100. It is, then, this twelve per cent which is the decrease in the ionizing-power of the mixed beam, produced by interposing ten cm. of lead; Rossi thinks that the data prove that the photons are less penetrating than the electrons.

If so, how can it be that the charged particles responsible directly for the ionization have so great a penetrating-power? The electrons expelled from atoms by X-rays, and by the gamma-rays of known wave-length emerging from radioactive substances, are definitely less penetrating than the photons which set them on their ways. No theory gives the slightest indication that this relation should eventually be reversed as the frequency and penetrating power of the gamma-rays

[19] I take equation (3) from Rossi's paper in the *Zeitschrift für Physik*, where the reader can also find equations for cases somewhat more complicated but probably not more plausible. The region beyond R is often called the region where "the primary beam is in equilibrium with its secondaries." Notice also the argument of Kulenkampff, mentioned in footnote 11. I am indebted to Dr. P. M. Morse for a discussion of these questions.

increases. True, the theories are based on the assumption that the energy of the ionizing particles is derived from the photons: if some of them originate from disintegrated atom-nuclei, as the seeming presence of protons among them suggests, the energy of these may be greater than that of the primary photon itself.

Could we dispense altogether with the notion of a beam of electromagnetic waves coming from above, and imagine that there are no cosmic rays other than these fast-flying particles, which then must be supposed to come down into the lower atmosphere from above? This possibility cannot be dismissed; against it, however, speaks the strong testimony of several observers who have travelled far and wide in the search for indications that the intensity of the rays varies from point to point on the surface of the earth. For the earth is a magnet; and while the strength of its field is minute compared with that prevailing within a few inches of an electromagnet or even an ordinary horseshoe magnet, the extent thereof is so great that electrons or protons coming up to our planet from interstellar space are liable to be enormously deflected. Charged particles of energy-values such as I have been mentioning (millions or tens of millions of equivalent volts) should reach the earth prevailingly in the region of the magnetic poles, if they start uniformly from all directions. Now, Bothe and Kolhörster took their instruments—three electroscopes and a pair of counting-tubes—on a cruise through Arctic waters; the ship started from Hamburg and returned there, its course encircling Iceland and passing close to Spitzbergen; there was no sign of a systematic variation of the readings. Millikan, whose data in California had agreed with his data in Bolivia in the admirable way which Fig. 4 displays, made a still finer test: in the summer of 1930 he went "to the settlement which is much the nearest to the earth's north magnetic pole of any settlement on earth, namely, Churchill, 750 miles due south of the pole on the west side of Hudson's Bay—at present a construction camp. . . . The mean results, when compared with those similarly taken at Pasadena during the last week in July and the first in August, show that the cosmic rays have precisely the same intensity at Churchill, in latitude 59, as at Pasadena in latitude 34, the mean results in the two places being 28.31 ions per cc. per sec. and 28.30 ions per cc. per sec. respectively, as measured in my particular electroscope. I think the error in these measurements cannot possibly be as much as 1 per cent." A British expedition to the Antarctic made measurements of cosmic rays within 250 miles of the south magnetic pole, and found the same intensity there as in Australia.[20] It is difficult to doubt that anything

[20] K. Grant, *Nature*, **127**, 924 (1931).

so constant must come from without the terrestrial world, and on its way be exempt from the influence of the earth's magnetic field.[21]

Of the origin of the cosmic rays I have not spoken in this article, having so chosen its title as to exclude that mighty subject from its scope. In the data and in the nature of these rays there is enough of the mysterious and enough of the extraordinary, to suffice for an introduction. Modest as the data seem, what they reveal is sensational. It is found that the atmosphere of this earth is being traversed by ionizing particles, of which the qualities are amazing. If they are such familiar corpuscles as protons or electrons, their energy must be of the order of tens of millions of equivalent volts, values without precedent in our experience. If they are neutral particles they are in themselves unprecedented. If they are electrons or protons which derive their energy from photons or corpuscles of light, these last must have energy greater, frequency higher and wave-length smaller, than any form of light of which we have previous knowledge. Such quantities of energy, be it remembered, are enormously greater than the largest which atoms can emit in the course of their normal lives; they are several times larger even than those which are emitted by collapsing atom-nuclei, in those processes of transmutation of which we already have knowledge. If they come from individual atoms they must arise from processes heretofore unknown, of transmutation or synthesis or annihilation. The constancy of the effect of the ionizing rays, its independence (be it absolute or only approximate) of weather and time and direction and the earth's magnetic field, implies that these processes are spread, not over the earth nor even over the solar system by itself, but throughout the whole of the cosmos.

Literature

The literature of cosmic-ray research has already swollen to such proportions, that a complete bibliography would cover several pages of this journal. There seems to be no book devoted altogether to the subject, though in treatises on atmospheric electricity—I name especially *Die elektrische Leitfähigkeit der Atmosphäre*, by V. F. Hess, translated as *The Electrical Conductivity of the Atmosphere and its Causes* (Van Nostrand, 1926)—a chapter or a section is usually assigned to it.

The major publications of Millikan's school have appeared in the *Physical Review*, as follows:

R. A. Millikan & I. S. Bowen, *Phys. Rev.* (2), **27**, 353–361 (1926) (sounding-balloon observations).

Millikan & R. M. Otis, *ibid.*, **27**, 645–658 (1926) (mountain-peak and aeroplane observations).

Millikan & G. H. Cameron, *ibid.*, **28**, 851–868 (1926) (lakes in California).

Millikan & Cameron, *ibid.*, **31**, 163–173 (1928) and 921–930 (lakes in California and Bolivia).

[21] Nevertheless there have been opposing data; Clay, on a voyage to Java, from the Mediterranean, found the ionization increasing as he approached the equator, and Corlin is said to have observed an increase from south to north along the Scandinavian peninsula; recall also Corlin's assertion mentioned on page 169.

Millikan & Cameron, *ibid.*, **32**, 533–557 (1928) (interpretation).
Millikan, *ibid.*, **36**, 1595–1603 (1930) (approach to north magnetic pole).
Millikan & Cameron, *ibid.*, **37**, 235–252 (1931) (most recent and complete ionization-*vs.*-depth curve).

To this list should be added Millikan's address before the B. A. Symposium on the Evolution of the Universe (*Nature*, **128**, 709–715; 24 Oct. 1931), and the account of C. D. Anderson's work probably soon to be published.

Regener's observations at various depths of water are described in *Naturwissenschaften*, **17**, 183–185 (1929), with a later and regrettably brief note in *Phys. ZS.*, **31**, 1018–1019 (1930). Piccard's observation at the altitude of 16 km. is mentioned in the *Bulletin* of the French Physical Society for 4 Dec. 1931. Swann's observations at three altitudes appear in *Jour. Franklin Inst.*, **209**, 151–200 (1930), and an observation by A. H. Compton in *Phys. Rev.* (2) **39**, 55 (1932).

Skobelzyn's experiments by the expansion-chamber method are published in *ZS. f. Phys.*, **54**, 686–702 (1929); see also P. Auger & D. Skobelzyn, *C. R.*, **189**, 55–57 (1929). L. M. Mott-Smith and G. L. Locher made simultaneous observations by this and the Geiger-counter method (*Phys. Rev.* (2), **38**, 1399–1408; 1931).

The absorption-experiments with Geiger counters cited in this article are due to W. Bothe & W. Kolhörster, *ZS. f. Phys.*, **56**, 751–778 (1929), and B. Rossi, *ibid.*, **68**, 64–84 (1931); see also Rossi, *Lincei Rendiconti* (6), **13**, 600–606 (1931).

Experiments with Geiger counters to seek for deflection of ionizing particles by magnetic fields: Mott-Smith, *Phys. Rev.* (2), **37**, 1001–1003 (1931); Rossi, *Lincei Rendiconti* (6), **11**, 478–482 (1930), and *Nature*, **128**, 300–301 (22 Aug. 1931); L. F. Curtiss, *Phys. Rev.* (2), **35**, 1433; (1930).

Search for influence of earth's magnetism: Millikan, *ll. cc.;* K. Grant, *Nature*, **127**, 924 (20 June 1931) (approach to south magnetic pole); Bothe & Kolhörster, *Berl. Ber.* (1930) (cruise in Arctic waters); J. Clay, *Proc. Amsterdam Acad.*, **30**, 1115–1127 (1927), and **31**, 1091–1097 (1928) (cruise to tropics); Rossi, *Nuovo Cimento*, **8**, 3–15 (1931); W. M. H. Schulze, *Nature*, **128**, 837–838 (14 Nov. 1931) (magnetic storms).

On the dependence of cosmic-ray ionization on time there have been many researches, mostly by people desirous of finding (or of *not* finding) influences of barometric pressure, weather, variations of the earth's magnetic field, the position of the sun, the position of the Milky Way or other remote bodies. Most if not all of the attempts to determine the absolute value of the ionization to be ascribed to cosmic rays, on or above the ground, have involved by necessity studies of fluctuations. Fig. 9 of this article comes from G. Hoffmann, *ZS. f. Phys.*, **69**, 703–718 (1931); his paper contains references to earlier German work, some of it in a geophysical journal. Fig. 10 above comes from G. von Salis, *ZS. f. Phys.*, **50**, 793–807 (1928). Millikan discusses the fluctuations in his 1930 paper. The references to Hess' estimate of the influence of the sun are: *Nature*, **127**, 10–11 (3 Jan. 1931) and *ZS. f. Phys.*, **71**, 171–178 (1931); earlier work is mentioned there.

The estimate made by Kolhörster and Tuwim of the average number of ion-pairs produced (in air under standard conditions) per centimetre path of one of the high-speed ionizing particles detected by the Geiger counters, is stated in *ZS. f. Phys.*, **73**, 130–136 (1931); for the notation it seems to be necessary to study an earlier paper in *Berliner Berichte*, 1931.

For effects associated with the "transition-zone" mentioned above on page 180, see E. Steinke, *Phys. ZS.*, **31**, 1019–1022 (1930) and the earlier work there cited; A. Corlin, *ibid.*, 1065–1071.

For the variation of the ionization inside a high-pressure air-chamber with the pressure of the air, see J. W. Broxon, *Phys. Rev.* (2), **37**, 1320–1337 (1931) and the literature there cited; E. G. Steinke & H. Schindler, *Naturwiss.* **20**, 15–16 (1932) and literature there cited. For the theory mentioned on page 170 *supra*, see Millikan & Bowen, *Nature*, **128**, 582–583 (1931); A. H. Compton, R. D. Bennett & J. C. Stearns, *Phys. Rev.* (2) **38**, 1565–1566 (1931); Steinke & Schindler, *loc. cit.*

Abstracts of Technical Articles from Bell System Sources

The Emission of Secondary Electrons from Tungsten.[1] A. J. AHEARN. An apparatus is described for investigating critical potentials in the emission of secondary electrons from tungsten. Measurements on the velocity distribution in the primary beam show that secondary electrons from the electron gun are absent. Tube characteristics which might introduce spurious critical potentials in the secondary emission from tungsten appear to be absent. By heat treating the tungsten and cleaning up residual gases, maxima and critical slope changes were developed below 40 volts. With sensitive methods of measuring and plotting the data, critical potentials within the range from 40 to 500 volts were observed only at the following uncorrected voltages: 70, 108, 208, 297 volts. All but the 70-volt effect disappeared eventually after heat treatments of the tungsten target. Thus when the tungsten surface is most free from contamination, critical potentials persisted only at the following uncorrected voltages: maxima at about 3.5 and 8 volts and slope increases at 24, 33 and 70 volts. The phenomena may be associated with the diffraction of electrons or the production and absorption of characteristic soft x-rays. Regardless of the mechanism operating at the critical potentials, their decrease or elimination beyond 40 volts points strongly to effects of surface contamination rather than to characteristics of tungsten.

Electrolytic Phenomena in Oxide Coated Filaments.[2] J. A. BECKER. A critical survey of the literature shows that the current through the oxides in oxide coated filaments is carried by electrons, negative oxygen ions, and positive barium ions. The proportion of current carried by each depends upon the exact composition and method of preparation of the oxide coating, on the heat treatment and on previous electrolytic effects. Presumably the conductivity is greatly affected by barium and oxygen dispersed through the oxide. New experimental results show:

1. For a particular BaO + SrO filament, the conductivity C was given by

$$1.71 \times 10^{4} \frac{1.73 \times 10^{4}}{T} + 5.55 \times 10^{-3} \frac{0.62 \times 10^{4}}{T}.$$

[1] *Phys. Rev.*, November 15, 1931.
[2] *Trans. Electrochem. Soc.*, Vol. LIX, 1931.

2. The current is proportional to the voltage only so long as the current is small; otherwise the products of electrolysis alter the conductivity.

3. Polarization currents are caused by the Ba and O which are produced by electrolysis. These currents decrease rapidly even at temperatures near 500° K., thus showing that Ba and O diffuse at low temperatures.

Some Observations of the Behavior of Earth Currents and their Correlation with Magnetic Disturbances and Radio Transmission.[3] ISABEL S. BEMIS. This paper presents correlations between the abnormal earth currents noted during magnetic storms and transoceanic radio transmission on both long and short waves. The radio transmission data were collected on the telephone circuits operating between New York and London and between New York and Buenos Aires. The earth current data were collected on two Bell System lines extending approximately a hundred miles north and west from New York. The results of this work establish facts which have been known in a general way for some time.

The direction of flow of abnormal earth currents in the neighborhood of New York seems to be along a northwest-southeast line. Coincident with such abnormal currents are periods of poor short-wave radio transmission. However, on long waves, daylight transmission over transatlantic distances is improved. On the short-wave circuit to Buenos Aires, transmission is adversely affected but only to a moderate extent.

The Propagation of Short Radio Waves over the North Atlantic.[4] C. R. BURROWS. Transmission conditions for each season are shown by "surfaces" giving the received field strength as a function of time of day and frequency. These show that frequencies near 18 mc are best for daytime transmission. In summer the best frequencies for nighttime transmission are those near 9 mc. In winter an additional frequency near 6 mc is required during the middle of the night. A frequency (such as 14 mc) intermediate between the day and night frequency is useful during the transition period between total daylight and total darkness over the path. Day-to-day variations change the periods of usefulness of these frequencies. In particular the period of usefulness on 14 mc sometimes extends so that it is the best daytime frequency.

Transmission conditions on undisturbed days were found to be the

[3] *Proc. I. R. E.*, November, 1931.
[4] *Proc. I. R. E.*, September, 1931.

same for the same time of year on different years. These undisturbed transmission conditions are presented by "normal" surfaces. Comparison of these surfaces shows that the higher frequencies are less attenuated in winter. Reception on the highest frequency, 27 mc was best in winter; in summer this frequency was never heard.

The effect of solar disturbances on short-wave transmission is to reduce reception on all frequencies. Sometimes the higher frequencies are the more adversely affected. Some of the possible causes of these disturbances are discussed.

From the measurements made on "static" at New Southgate, data on the variation of its field strength as a function of frequency, time of day, and season are given.

Methods of High Temperature Treatment.[5] PAUL P. CIOFFI. The object of this paper is to describe methods which have been developed in this laboratory for treating metals, chiefly iron and its alloys, at all temperatures up to about 1700° C., considerably above the melting point of iron, and in any atmosphere ranging in pressure from 10^{-8} mm. of mercury to 20 or more atmospheres. The methods of heating, and the forms of the materials heated fall into three rather well-defined groups: (1) Long wires and tapes heated to any temperature up to a few degrees below the melting point by passing currents through them in the presence of a gas (*a*) under a pressure of one atmosphere or less, or (*b*) under a pressure up to 20 atmospheres. (2) Toroids heated by induction in a gas atmosphere, with the gas pressure ranging from 10^{-8} mm. to one atmosphere. (3) Any shape of specimen heated up to about 1700° C. in a molybdenum wound furnace in a gas pressure ranging from 10^{-3} mm. to 20 atmospheres.

In investigations dealing with the effect of heat treatment on the properties of materials, the temperature range covered, often limited by the lack of suitable facilities, is likely to obscure important or interesting effects produced by heat treatments at high temperatures. The new high values of the magnetic permeability of iron recently reported were made possible by employing the methods here described; in fact, these methods were all developed in connection with the investigation of the magnetic properties of materials as dependent upon temperature and time of heat treatment and rate of cooling in any atmosphere. These methods are also applicable to other metals which can be shaped into long wires or tapes or toroids.

[5] *Jour. Franklin Institute*, November, 1931.

Methods for Measuring Interfering Noises.[6] LLOYD ESPENSCHIED. This paper outlines various methods of measuring interference, particularly in radio telephony, which have been found useful in the Bell System.

Reverberation Time Measurements in Coupled Rooms.[7] CARL F. EYRING. The decay of residual sound in simple reverberant enclosures has been the subject of much study. Out of these investigations along with other important developments formulæ have emerged which may be used to calculate the reverberation time of enclosed rooms, the one proposed by Sabine being applicable to live rooms, but a more general one being necessary if applied to dead rooms. These formulæ may be used in the study of simple enclosures in which the sound is diffuse, the absorbing material is well distributed, and the decay of the residual sound is exponential in time. They may not be applied indiscriminately to complex structures for the author has shown that the curves illustrating the decay of the residual sound intensity level may not be straight under certain conditions for coupled rooms of different natural reverberation times or even for a single room with no sound diffusing scheme and non-uniformly distributed absorption.

The present paper presents further data on acoustically coupled rooms and offers a theoretical study and formulæ applicable to such complex structures. Somewhat similar studies have been made by Buckingham and Davis in the investigation of sound transmission through partitions. But before the coupled room formulæ which are based on certain idealizations can be applied in general to auditoriums with various types of balconies and under-balcony spaces, they must be carefully checked by a thorough experimental study of typical theatres. Thus it is expected that empirical formulæ, based on theory and experiment, may be evolved for each type of complex structure.

Recently developed instrumental methods of measuring reverberation time, especially those methods that measure the decay history of the residual sound, give promise of being the tools needed in this study. The results recorded in this paper were made on a meter described by Wente and Bedell in their paper "Chronographic Method of Measuring Reverberation Time" and by the author. This instrument plots almost automatically the intensity level of the residual sound measured in db and the time.

[6] *Proc. I. R. E.*, November, 1931.
[7] *Jour. Acous. Soc. Amer.*, October, 1931.

When the sound source is cut off there will not only be a decay of sound due to the absorption of the walls, fixtures, etc., but the interference pattern will continually shift causing the actual rate of decay of the sound intensity level at a point to fluctuate about the rate of decay caused only by absorption. This fluctuation may be very pronounced and actually changes from point to point in the room, but may be minimized by the use of a warble tone and if necessary by the proper placement of the transmitter. With these precautions, and they are used in all the reverberation time measurements presented in this paper, the effect of interference is greatly reduced. The average rate or rates of decay δ (db per second) of the sound intensity level can at once be obtained from the slope of the straight line or lines which best fit the series of experimental points. The reverberation times T (time per 60 db) used in this paper have been calculated using the relation

$$T = \frac{60}{\delta}.$$

The Effect of Exposure and Development on the Quality of Variable Width Photographic Sound Recording.[8] DONALD FOSTER. This paper deals with the dependence of the quality of variable width recording on the conditions of exposure and development. When the widths of the images employed in recording or reproducing are comparable with the wave-length of the record the exposed portion of the record is not uniformly exposed. The record is attenuated in amplitude as the frequency is increased, and harmonics are introduced whose relative intensities depend on the contrast of development and on the frequency. When the exposure of the record occupies the linear range of the H & D curve, and when the product of the gammas of the negative and the positive is equal to unity, the record is practically free from spurious harmonics. The amount of non-linear distortion is calculated for the case when the over-all gamma is equal to two; and it is shown that Cook's analysis of the aperture effect gives a superior limit to the distortion obtainable by overexposure or by over or underdevelopment. The effect of the unavoidable non-uniform illumination of the images is considered.

The Vectorial Photoelectric Effect in Thin Films of Alkali Metals.[9] HERBERT E. IVES. It is assumed that the photoelectric effect exhibited by thin films of alkali metals on specular platinum surfaces is proportional at any wave-length to the electric intensity just above

[8] *Jour. S. M. P. E.*, November, 1931.
[9] *Phys. Rev.*. September 15, 1931.

the platinum. This electric intensity is found, using the optical constants of platinum, by computing the intensities of the wave patterns formed by the interference of the reflected and incident beams. These computations are made for various angles of incidence and for light polarized in and at right angles to the plane of incidence. The intensities thus found exhibit very large ratios of value for the two planes of polarization, in striking agreement with the characteristics of the vectorial photoelectric effect. The changes of amplitude of the perpendicular electric vector on entering the alkali metal film, as computed from the optical constants of the alkali metal, account for the experimentally found low values of the emission ratios at long, and their high values, at short wave-lengths.

The Photoelectric Effect from Thin Films of Alkali Metal on Silver.[10] HERBERT E. IVES and H. B. BRIGGS. The thin films of alkali metals which spontaneously deposit in vacuo on other metals have long been known to exhibit photoelectric effects which vary in amount and character, depending on the underlying material, but the exact nature of this dependence has been obscure. Silver, because of its region of exceedingly low reflecting power in the ultraviolet and the accompanying variation of optical constants, is exceptionally well suited for studying the influence of the underlying metal. It is found that the region of low reflecting power profoundly affects the photoemission, but in a manner not to be explained simply by reduction of light reflected back through the alkali metal film or by the absorption of light by the silver. The results obtained are very satisfactorily explained upon computing, from the optical constants, the intensity at the surface, of the interference pattern formed by reflection just above the silver surface. The positions of the maxima and minima of photoemission, and their variations with angle of illumination and plane of polarization are accurately indicated.

The Applicability of Photoelectric Cells to Colorimetry.[11] HERBERT E. IVES and E. F. KINGSBURY. It is the purpose of this paper to consider critically the requirements for a precision physical colorimeter, and to estimate, in the light of a large body of experimental data on the new types of photoelectric cells, to what degree the requirements for physical colorimetry may be met at the present time. The paper is intended to be very specifically limited to the problems of precision color measurement, and it is assumed that the reader is already acquainted with the principal facts relating to photoelectric cells and

[10] *Phys. Rev.*, October 15, 1931.
[11] *Jour. Optical Soc. America*, September, 1931.

their use in ordinary photometric measurements, in which no pronounced color differences are faced. Our study does not concern itself with certain problems of sorting and selection, commonly spoken of as "color" measurement, in which the nature of the colors concerned and their range of deviation from certain standards are known in advance, thereby greatly simplifying the problem.

A Moving Coil Microphone for High Quality Sound Reproduction.[12] W. C. Jones and L. W. Giles. A microphone is described in this paper which retains all of the inherent advantages of the moving-coil type of structure but unlike the earlier forms of this microphone responds uniformly to a wide range of frequencies. It is more efficient than the conventional form of condenser microphone and its transmission characteristics are unaffected by the changes in temperature, humidity and barometric pressure encountered in its use. Unlike the condenser microphone the moving-coil microphone may be set up at a distance from the associated amplifier and efficient operation obtained. Owing to its higher efficiency and lower impedance it is less subject to interference from nearby circuits. It is of rugged construction and when used in exposed positions is less subject to wind noise.

The Shot Effect in Photoelectric Currents.[13] B. A. Kingsbury. The shot effect, as it occurs in a photoelectric current, has been used to secure an evaluation of the electron charge. A new and original method of amplifier calibration, which involved the use of a modulated light beam, simplified the measurements and the computation of the result. In the absence of space charge, the experimental value of the electron charge was 1.61×10^{-19} coulombs for a thermionic current, and about 25 per cent greater for a photoelectric current. It was found that the shot effect is enormously increased in photoelectric currents which are amplified by collision ionization. Statistical variations which might be expected to occur in a beam of radiant energy could not be detected, since, within the limits of experimental accuracy, the shot effect in photoelectric currents was found to be independent of the frequency of the light producing electron emission.

Some Acoustical Problems of Sound Picture Engineering.[14] W. A. MacNair. The purpose of this paper is to point out that many advances in acoustical engineering have been necessary in order to understand and control adequately the conditions under which modern

[12] *Projection Engineering*, October, 1931.
[13] *Phys. Rev.*, October 15, 1931.
[14] *Proc. I. R. E.*, September, 1931.

sound pictures are recorded and reproduced. To illustrate this point, some of the acoustical problems encountered at Bell Telephone Laboratories are discussed. The sudden and successive changes in sound intensity level to be expected in a room during the growth and decay of sound from an intermittent source are pointed out. The necessity of using the more general reverberation time formula, which was developed over a year ago, when dealing with comparatively "dead" rooms, is indicated. One type of acoustical distortion which is due to interference is discussed together with the measures necessary to minimize it in sound pick-up work. These phases of acoustical engineering have been selected for discussion from many which confront the engineer in this field.

An Interpretation of the Selective Photoelectric Effect from Two-Component Cathodes.[15] A. R. OLPIN. Evidence is produced to support the view that photoelectrically selective, two component cathodic surfaces are crystalline in nature. Then, assuming that Fowler's equation for the energy of electrons selectively transmitted through a single potential valley $[W = (n^2h^2/8md^2)]$ is equally valid for the energy of electrons selectively transmitted through the periodic sequence of valleys characteristic of the potential field within a crystal, and that all of the energy of photoelectrons is acquired from the incident light quanta, the wave-lengths of light to which such a surface should respond selectively can be computed. Such computations have been made with d equal to the internuclear distance between electro-positive ions in the lattice structure of alkali metal hydride, oxide and sulphide crystals. The hydride crystals belong to the sodium chloride type and the oxide and sulphide crystals are supposedly of the calcium-fluoride type. The correlation between these computed values and the positions of the observed selective maxima is exceptionally good. Moreover, the fact that the alkali metal hydrides exhibit but one selective maximum and the oxides two or three maxima is in keeping with the geometry of their respective crystalline types.

Some Physical Concepts in Theories of Plastic Flow.[16] R. L. PEEK, JR., and D. A. McLEAN. A review is given of the considerations involved in the development of theories of deformation applicable to the flow of soft solids in capillary tubes and under similar steady state conditions. It is pointed out that the limitation to special test conditions (particularly to steady states) makes it impossible to

[15] *Phys. Rev.*, November 1, 1931.
[16] *Jour. Rheology*, October, 1931.

distinguish in all cases between the results of different physical hypotheses as to the mechanism of flow resistance. A review is given of the physical distinctions which can be determined by such experiments. It is noted that the criterion that the results from different capillary tubes can be expressed in the form $Q/R^3 = F(PR/2L)$ serves to distinguish those types of flow in which the resistance is dependent only on the rate and not on the extent of the deformation, except when slip occurs. This last can be distinguished from quasi-laminar flow (in which the resistance depends on the amount of deformation) by tests made with capillaries of common radius but of different lengths.

The other important physical distinction that can be observed is between those materials which show a yield value and those that do not. For those that do, a new type of equation is obtained, of which the Bingham-Buckingham equation is a special case. In this new general form the relation postulates an initial shear stress which must be exceeded before flow takes place, and a lower constant stress which is effective in opposing flow once the latter has commenced.

For cases of flow in which no yield value is observed, another new form is given which is based on the physical concept, common to most theories of such flow, of an effective viscosity varying with the stress intensity between upper and lower limits. It is shown that the character of this relation is similar to that given by the empirical formula $Q/R^3 = K\left(\frac{PR}{2L}\right)^n$. The use of this equation in expressing and interpreting experimental results is discussed.

Intercontinental Radiotelephone Service from the United States.[17] J. J. Pilliod. Radiotelephone service between the United States and Europe was established January 7, 1927 with one circuit and with service to limited areas. Facilities and service have been greatly improved and extended and rates have been reduced. Present scope of service is described and reference made to consistent increases in transatlantic telephone messages handled. This increase indicates that this service is being found of increasing value by the public.

Extent of ship-to-shore radiotelephone service from the United States is outlined. Arrangements for service to Buenos Aires and Rio de Janeiro are described, these differing from arrangements used for service to Europe in that operation to these two cities was planned on a part time basis. Proposed short-wave system for operation with

[17] Presented at Pacific Coast Convention of A. I. E. E., Lake Tahoe, Calif., August, 1931. Published in abridged form in Elec. Engg., September, 1931.

Bermuda and proposed new long-wave system to supplement existing facilities to Europe are mentioned.

A description of the new radiotelephone transmitting and receiving stations now being erected at Dixon and Point Reyes, Calif., respectively, is given. These stations will be connected to a terminal office at San Francisco and the system used for the establishment of radiotelephone service to the Hawaiian Islands and later on, to other transpacific points as may be required.

High-Frequency Atmospheric Noise.[18] R. K. POTTER. A method which has been employed in the measurement of high-frequency atmospheric noise is described. Using this method measurements of noise over the range from 5 to 20 megacycles made in different parts of the United States and at different times of the year, show a distinct diurnal change in intensity similar to that for long-range high-frequency signal transmission. Except during periods of severe local disturbance noise on the lower frequencies is high during the night while on the higher frequencies the maximum occurs during the day. Simultaneous observation of crashes on different frequencies also suggests that the received atmospherics are largely transmitted by overhead paths. The variation in high-frequency atmospheric noise intensity during the passage of local electrical disturbance centers is shown. It is suggested that the intensity of atmospheric noise generated by these centers of electrical disturbance is inversely proportional to frequency. Measurement data are included showing the effect of sunrise and sunset, an eclipse of the sun, and disturbances in the earth's magnetic field upon the intensity of high-frequency atmospheric noise. Diurnal characteristics of high-frequency atmospheric noise on directive antennas facing England and South America and the noise reduction obtained by these arrays are illustrated. The possible location of distant centralized noise sources is discussed briefly.

The Grounded Condenser Antenna Radiation Formula.[19] W. HOWARD WISE. Exact formulas for the wave function and vertical electric field at the surface of the ground are derived for a vertical dipole of zero height.

[18] *Proc. I. R. E.*, October, 1931.
[19] *Proc. I. R. E.*, September, 1931.

Contributors to this Issue

O. B. Blackwell, B.S. in Electrical Engineering, Massachusetts Institute of Technology. After graduation, he entered the Engineering Department of the American Telephone and Telegraph Company and in 1919 was made Transmission Development Engineer. Mr. Blackwell has general supervision of transmission developments and has been prominently associated with progress in long distance wire and radio telephony.

Karl K. Darrow, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917; Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences. Some of his earlier articles on Contemporary Physics form the nucleus of a book entitled "Introduction to Contemporary Physics."

Clifford E. Fay, B.S. in Electrical Engineering, Washington University, 1925; M.S., 1927. Bell Telephone Laboratories, 1927–. Mr. Fay has been engaged principally in the development of power vacuum tubes for radio purposes.

J. B. Johnson, B.S., University of North Dakota, 1913; M.S., University of North Dakota, 1914; Ph.D., Yale, 1917; Engineering Department, Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. As Research Physicist, Dr. Johnson has been engaged in the development and application of special vacuum tubes and gaseous discharge devices.

Frederick B. Llewellyn, M.E., Stevens Institute of Technology, 1922; Ph.D., Columbia University, 1928. Western Electric Company, 1923–25; Bell Telephone Laboratories, 1925–. Dr. Llewellyn has been engaged in the investigation of special problems connected with radio.

D. A. McLean, B.S. in Chemical Engineering, University of Colorado, 1929. Bell Telephone Laboratories, 1929–. Mr. McLean's work in the Chemical Research Department has dealt principally with plasticity and viscosity studies of various materials.

H. NYQUIST, B.S. in Electrical Engineering, North Dakota, 1914; M.S., North Dakota, 1915; Ph.D., Yale, 1917. Engineering Department, American Telephone and Telegraph Company, 1917–19; Department of Development and Research, 1919–. Dr. Nyquist has been engaged in transmission work particularly relating to toll cables.

R. L. PEEK, JR., Columbia University, A.B., 1921; Met. E., 1923. Western Electric Company, Engineering Department, 1924–25; Bell Telephone Laboratories, 1925–. Mr. Peek is in the Materials Testing Division of the Apparatus Development Department.

E. E. SCHUMACHER, B.S., University of Michigan; Research Assistant in Chemistry, 1916–18. Engineering Department, Western Electric Company, 1918–25; Bell Telephone Laboratories, 1925–. Mr. Schumacher's work since coming with the Bell System has related largely to research studies on metals and alloys.