The Bell System Technical Journal

April, 1927

Developments in the Manufacture of Copper Wire[1]

By JOHN R. SHEA and SAMUEL McMULLAN

SYNOPSIS: This paper describes a new copper rod and wire mill located at the Western Electric Company's Plant at Chicago. It includes a brief survey of the copper rolling and wire drawing art at the time the investigation was started; a summary of tests made in varying the practice in rod rolling and wire drawing; and an outline of the work done by the Western Electric Company engineers in developing and designing new types of wire drawing machinery.

The rod mill is converting 225 pound wire bars into ¼″ rod in fourteen instead of the usual eighteen passes. This is accomplished by making heavier reductions in the first four passes while the copper is hot. The new wire mill incorporates many novel features, and the wire drawing machines are more compact in design and of considerably higher speed than those in general use.

The design of the wire mill was undertaken following a comprehensive survey of wire drawing processes and equipment used in this country and abroad. Part of this survey consisted of a study of the manufacture of diamond dies, it having been found that dies suitable for high speed drawing required a differently shaped "approach," a better polish, and a shorter "land," than those which were available for low speed work. The economies in floor space and plant investment due to the use of more compact and higher speed machinery are outlined. Some of the outstanding features in plant arrangement which contribute to more efficient operation are discussed in the concluding pages.

RAPID growth in the various branches of electrical communication accompanied by widespread research are constantly leading to the more efficient and economical meeting of the increasing demands for service. In this connection, one of the more recent and very interesting investigations indicated the possibility of effecting substantial improvement in the process of manufacturing copper wire. Accordingly, a comprehensive study of all the factors concerned was undertaken which resulted in the construction of a rod and wire mill at Chicago embodying many unique and improved features. A schematic layout is given in Fig. 1.

At the outset the sources of copper and its transportation were studied and it was found more economical to ship wire bars to Chicago for conversion into wire than to locate a wire mill near some of the large refineries and ship wire to the factory. It was also considered that this plan would reduce the investment in wire during the process of manufacturing cable and telephone apparatus.

[1] Read before the Midwinter Convention of the A. I. E. E., New York City, Feb. 8, 1927.

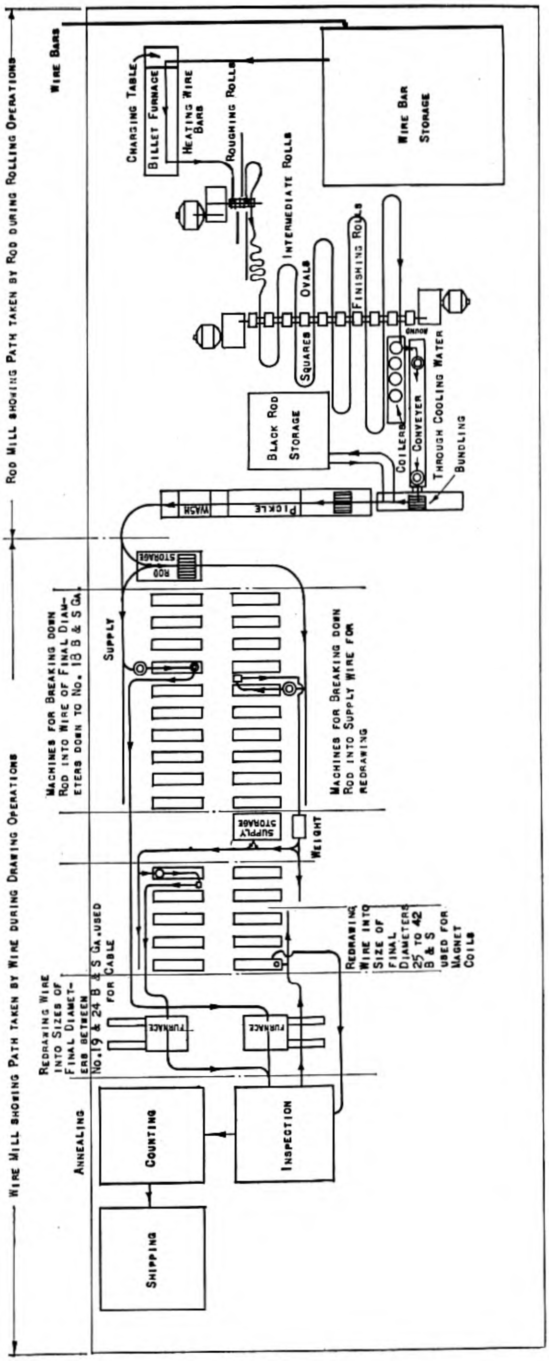


Fig. 1—Schematic layout of Western Electric Company's copper rod and wire mill at Chicago, Ill.

Rod Rolling Mill

The Rod Rolling Mill equipment consists of a billet heating furnace, a roughing mill, an intermediate mill, a finishing mill, coilers, conveyors, and pickling tubs. The mills are water-cooled and equipped with a down-draft exhaust which carries the fumes produced during the rolling operation to an air washer where the copper dust is recovered before the air is discharged.

The 225 pound wire bars as received in cars from the refineries are unloaded onto skids in the train shed and transported by an electric truck to the charging end of the billet heating furnace. Here they are transferred in groups of six by a hoist to the charging table, where a compressed air-pusher moves them along through the furnace which holds 120 bars. The bars are brought up to the required temperature for rolling as they move through the furnace, which is heated by fuel oil. When the bars reach the opposite end of the furnace they are withdrawn at about 1600° F. with a pair of tongs through the discharge door and pushed into the roughing mill one at a time. These tongs operate on a trolley suspended from a beam, which is in line with the first groove of the mill.

The roughing mill consists of three motor-driven rolls, one above the other. The bar, after passing through the first groove between the top and middle roll, drops upon feed rolls set in the floor and is returned through the second groove, between the middle and bottom roll; then raised into position and passed through the third groove, which is in the same rolls as the first pass. Five passes are made in this manner until its cross-section is reduced sufficiently for it to enter the intermediate mill. As the bar enters the roughing mill it is 54 inches long and about 4 inches square. When it leaves this mill it has been rolled into an oval cross-section and is about 124 feet in length. Formerly the last pass on this mill was handled manually, and recently a mechanical repeater has been added as illustrated by Fig. 2.

From the roughing mill the bar goes to the intermediate mill and is passed through the first pair of rolls. As it emerges an operator catches the end with a pair of tongs and passes it back through the next pair of rolls. The increased length between each pass at the intermediate and finishing mills is allowed to run out in a loop on a sloping iron covered floor on each side of the rolls. This catching and returning is repeated at each set of rolls until the original copper bar finally emerges a round, quarter-inch rod about 1200 feet long. This last pass goes through a guide pipe into a coiler, Fig. 3. The reductions in cross-section are illustrated in Fig. 4.

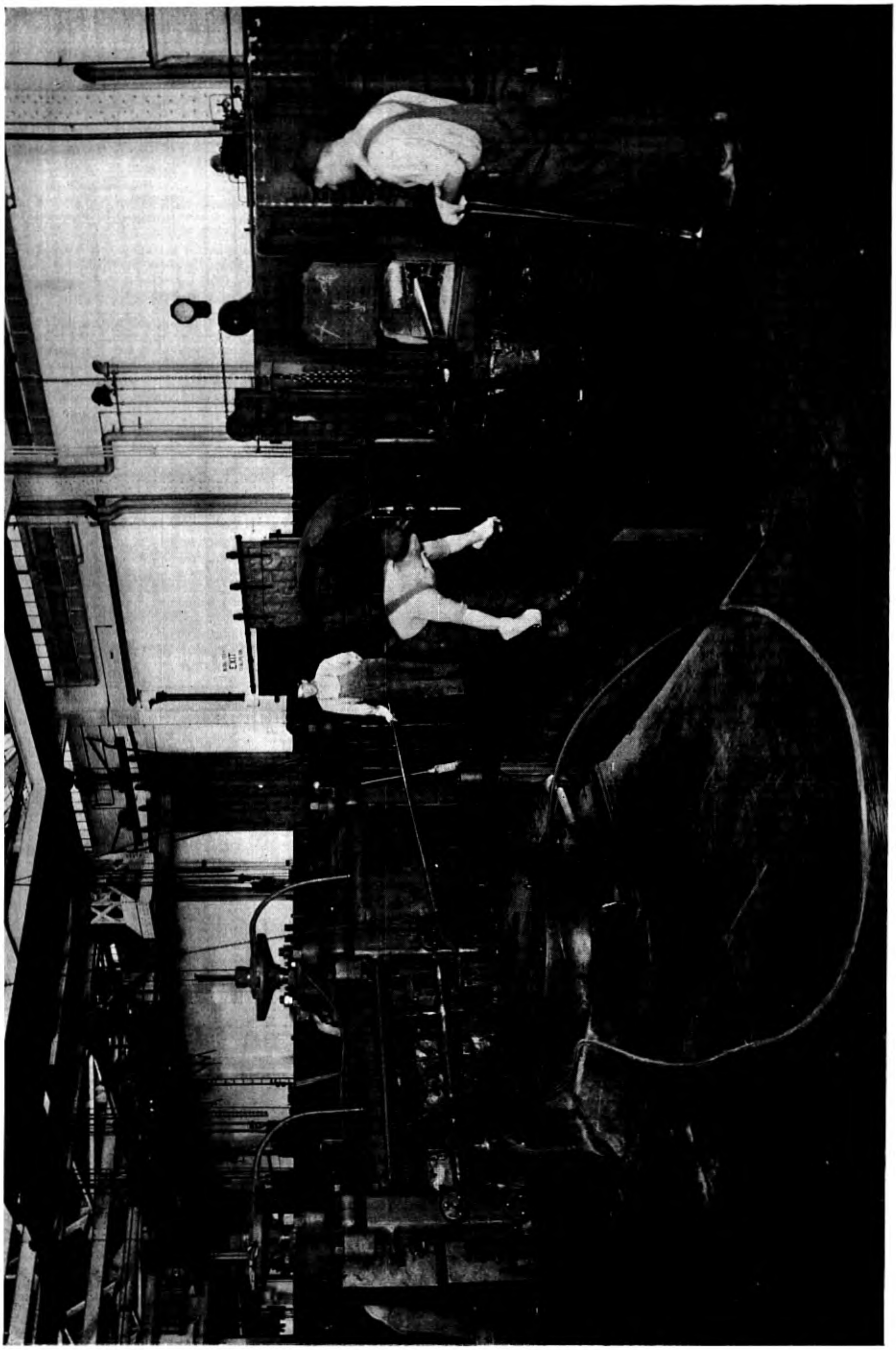


Fig. 2—View of roughing mill showing repeater on last pass

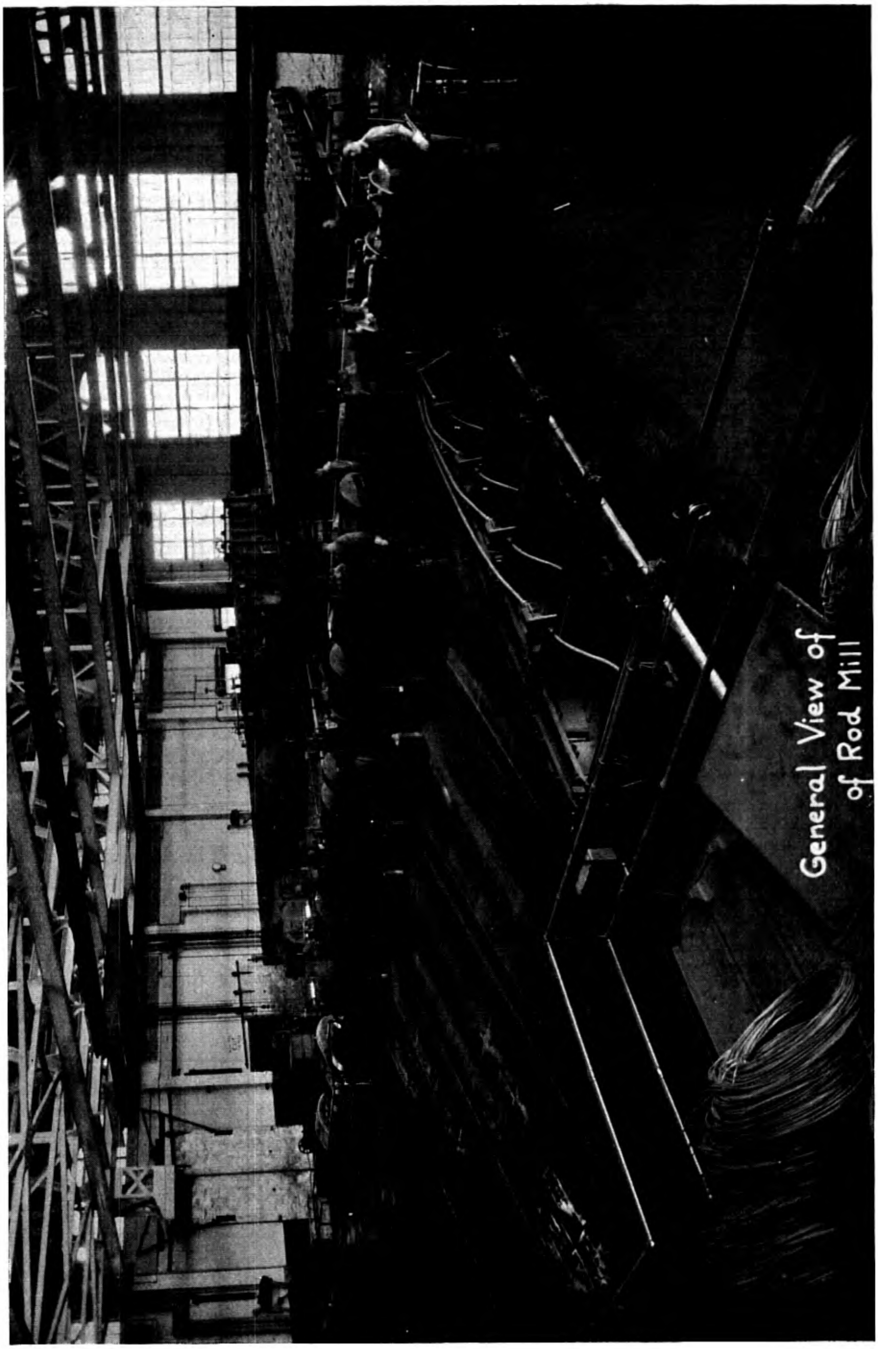


Fig. 3—View of intermediate and finishing mills and coilers

The coils are automatically unloaded from the coilers onto a conveyer, which carries them through cooling water in a tank underneath the floor. An appreciable amount of copper oxide scale is carried off

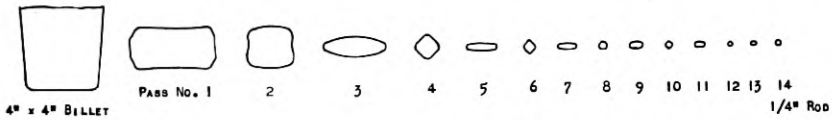


Fig. 4—Rod mill reductions—4 in. by 4 in. billet to ¼ in. rod

with the cooling water, and deposited in a reservoir from which it is later salvaged. Eighty-two seconds after entering the roughing mill the bar is a coil of ¼ in. rod ready to proceed on its way to the pickling tanks. The mill has a capacity of 70,000,000 pounds annually on a 48 hour per week basis.

While the diagram and photograph of the intermediate and finishing mills indicate for simplicity that the rod follows only a single path, in actual operation sufficient material is kept in the mill to practically

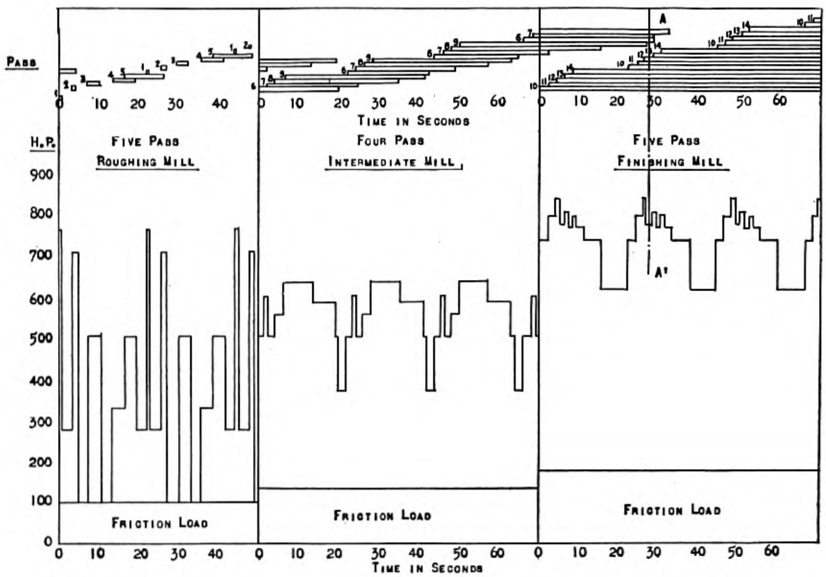


Fig. 5

maintain at least two rods in the finishing mill. This is illustrated graphically by that part of Fig. 5 which covers the finishing mill. Referring to line (A–A'), 11 reductions are being made in this mill at the same time, two for each of the first four pair of rolls and three on the final rolls. At this period in the cycle of operation 800 H.P. is required.

When the rod mill was started eighteen passes were in use by several of the most modern mills. A sixteen pass arrangement was adopted for the new mill, in which the metal was subjected to a greater amount of working in the earlier passes when it was hot. Later, as a result of further study, fourteen passes were adopted. Fig. 6 illustrates graphi-

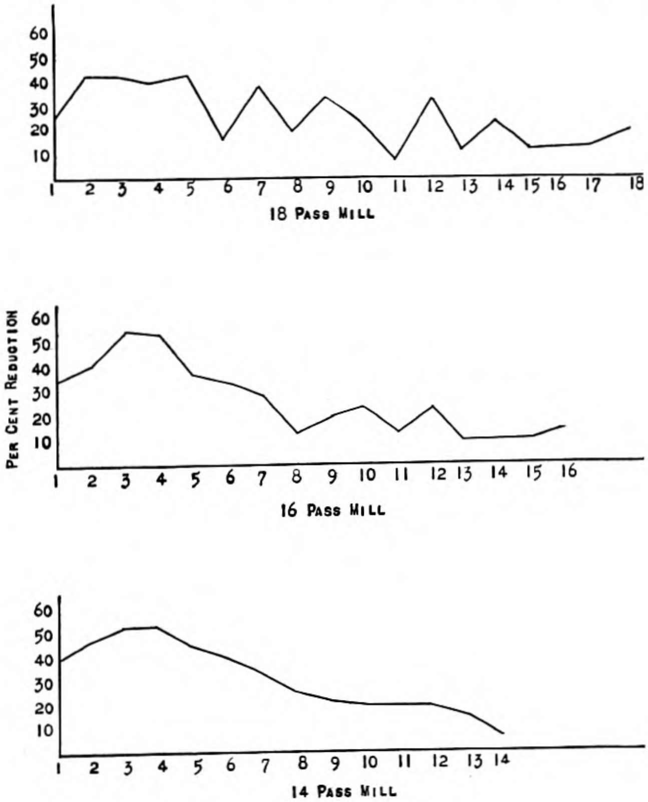


Fig. 6—Per cent reduction

cally the per cent reduction effected at each of the above passes. The reductions plotted as the abscissa are in terms of reduced area in cross-section at each pass and the passes reading from left to right are plotted as ordinates.

It is obvious that careful planning must be done in changing the number of passes in a mill in order not to exceed the safe working capacity of the mill rolls and stands. Such calculations have been made using rolling mill formulæ.[1] Based on the design of the mill using the eighteen pass arrangement the first four passes would operate at about 62, 100, 105, and 90 per cent of the safe working load of the mill. These same passes calculated on the basis of the sixteen and fourteen pass arrangement of the more sturdy mill at the Chicago plant operate at 86, 87, 90, 85 and 96, 96, 90, 90 respectively. This indicates that a further reduction may be made in the number of passes in the mill provided roll adjustment is improved.

Relation between Working and Physical Properties

It has been often stated that the more passes (i.e. the more gradual working) given the copper, the better the physical qualities of the rod. Actual tests (see Table I) made on representative lots of ¼ in. rod fail to confirm this impression.

TABLE I

Lot	Number of Passes	Elongation	Tensile Strength Lbs. per Square Inch
1	18	35.8%	33,752
2	18	40.0%	31,445
3	16	37.1%	32,468
4	16	41.0%	32,160
5	14	42.0%	32,391
	Average of 5 lots	39.5%	32,243

The averages indicate that fourteen pass rod is superior in elongation, and better than the total average in tensile strength.

Cleaning of Rod

When the coils emerge from the tank through which the rod coiler apron conveyor passes, they are cool enough to handle and after being tied with wire, several are lifted together by a monorail crane, and placed for thirty minutes in a pickling tank containing from 5 to 10 per cent free sulphuric acid, in order to remove the black oxide caused by oxidation of the hot copper in the air during rolling. The solution is maintained at approximately 120° F., and the copper content varies from 1 to 3 grams per 100 c.c. Experiments have shown a difference of less than 10 per cent in pickling time between the minimum and maximum acid used, the greater solubility being obtained from the weak solution. Actual results obtained were checked with Sidell's

[1] "Pass Limitation in Rolling Mill Practice," *Machinery*, July, 1918. "The Theory and Practice of Rolling Steel," Wilhelm Tafel.

Table of Solubilities (see Fig. 7). While a variation from the minimum to maximum acid concentration does not materially affect the pickling

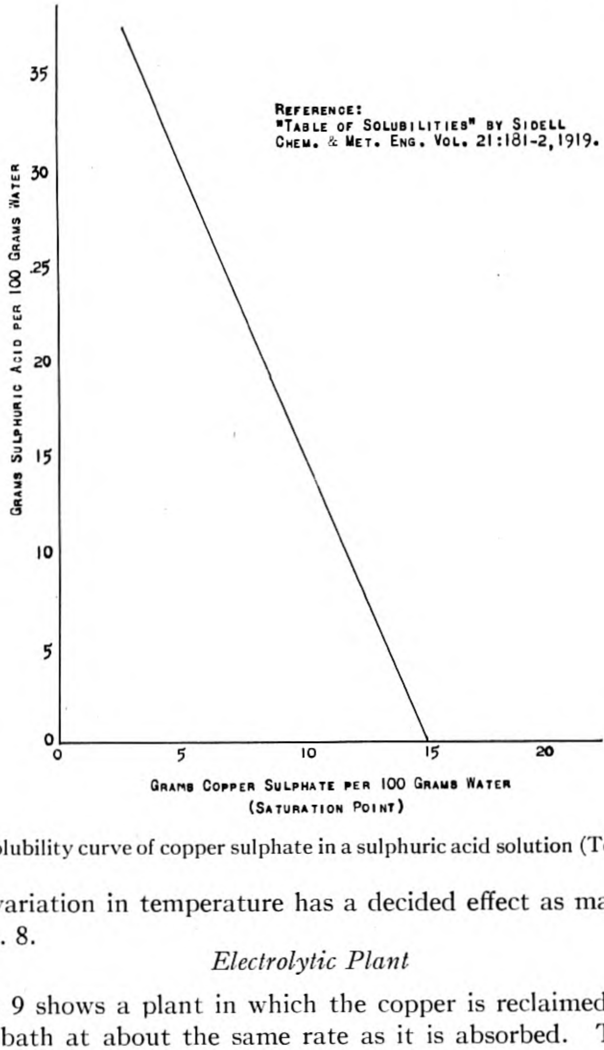


Fig. 7—Solubility curve of copper sulphate in a sulphuric acid solution (Temp. 25° C.)

time, a variation in temperature has a decided effect as may be seen from Fig. 8.

Electrolytic Plant

Figure 9 shows a plant in which the copper is reclaimed from the pickling bath at about the same rate as it is absorbed. This is accomplished by electrolytic deposition according to principles worked out and practiced in the large refineries which produce electrolytic copper.

The electrolytic system operates best with a minimum content of about 1 per cent copper and 5 per cent acid and a maximum of 3 per cent copper and 10 per cent acid. The copper and acid contents are kept as low as practicable to minimize "carrying out losses" [2] during the pickling operation. About 775 pounds of acid and 430 pounds of copper are recovered per day from the electrolyte. The anodes are operated at a current density of 5 amperes per square foot with a rate of deposition of about .00261 pound of copper per ampere hour.

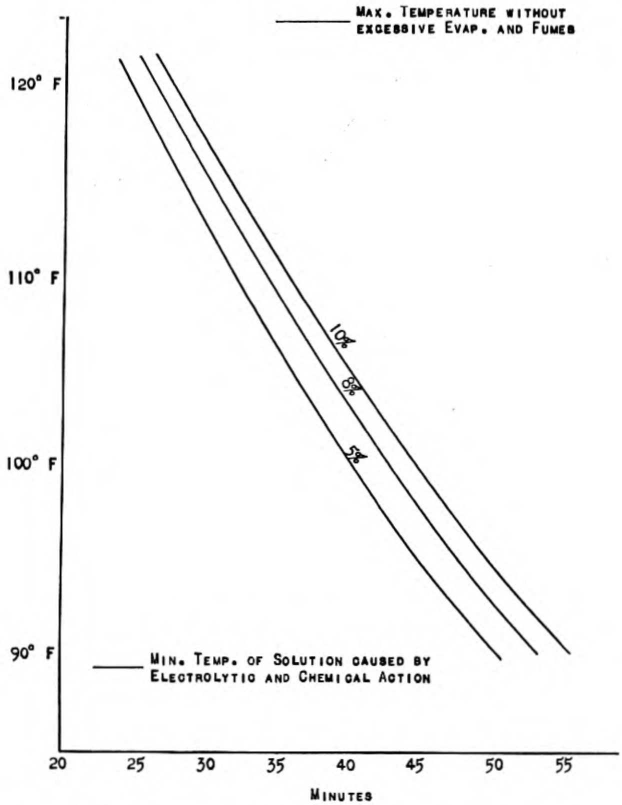


Fig. 8—Rate of pickling at practical acid concentrations

The heat generated in the plating tanks under normal operating conditions maintains a minimum temperature of about 90° F., throughout the acid system, and the maximum temperature is obtained through steam heating coils in the pickle tanks. Faster pickling would result from the use of higher temperatures but experience has shown that the additional steam and gas released above 120° F. results in unsatisfactory operating conditions.

[2] Pickling solution carried out when coils are removed from tank.

The coils of rod after pickling are thoroughly washed with lake water [3] at a pressure of about 70 pounds per square inch to remove loose

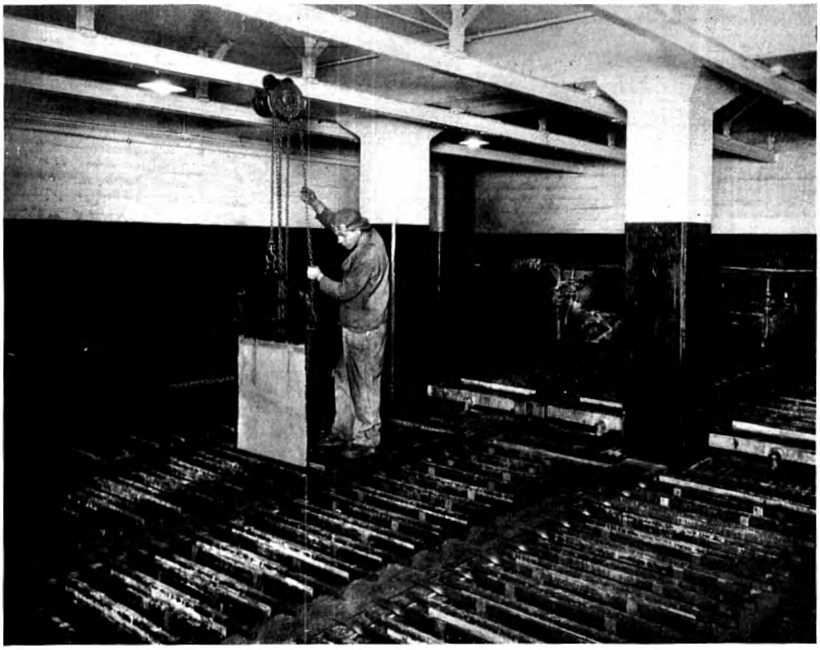


Fig. 9—Electrolytic recovery of copper from rod mill pickling solution

copper dust and acid, and then immersed in an alkaline fat solution to neutralize any trace of acid and to provide a protective coating against oxidation until converted into wire.

Wire Mill

The coils, after being pickled and washed, are carried by monorail cranes to the wire drawing machines where they are converted into wire of the desired size. The dies used in the heavy wire drawing machines are pulled into place at the starting end of the coil of rod on a die stringing machine (Fig. 10). The coil, with dies strung into position, is then placed in a heavy wire drawing machine.

The heavy gauges of wire, such as line wire, are drawn with one set-up on this machine; medium sizes, used in lead covered cable, are made by taking the wire as it comes from the heavy machine and re-

[3] Lake water is relatively free from mineral salts which would corrode the rod and affect the wire drawing compound.

drawing it on the intermediate machine; and finer sizes, commonly known as magnet wire, are produced by redrawing intermediate sizes.

The present capacity of the wire mill is approximately 42,000,000 pounds annually, and the sizes range from .165 in. line wire to 42 B. & S. (.00247 in.) gauge magnet wire. Provisions have been made in the construction of the building and its foundations so that the mill may be expanded in capacity when needed.

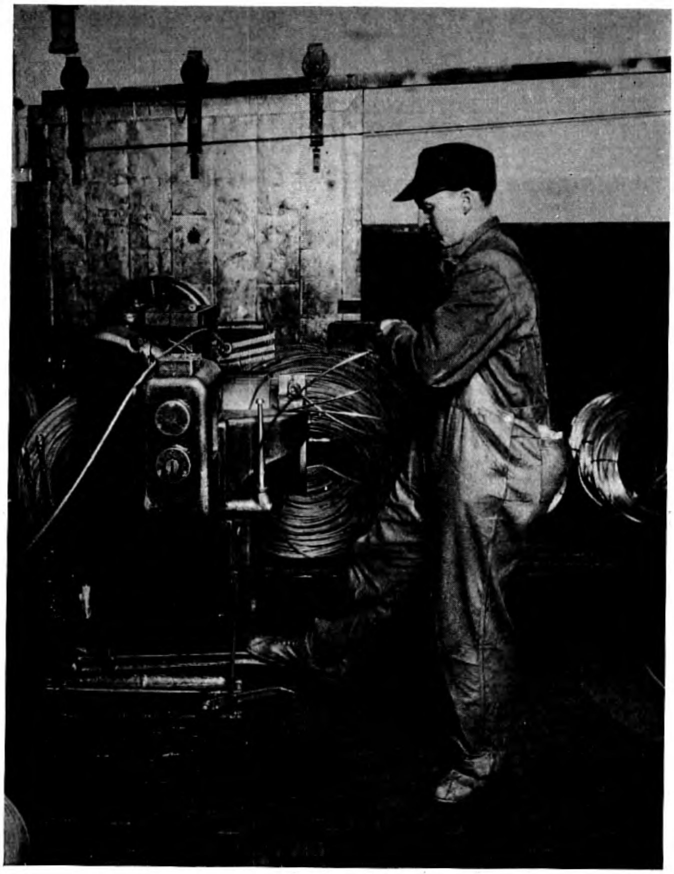


Fig. 10—Heavy wire die stringer

The No. 1 or heavy wire drawing machine shown by Figs. 11 and 12 draws line wire, heavy toll cable sizes, and supply wire for the loop cable wire machines. This ten die machine with its auxiliary equipment and operating area occupies a floor space of 270 square feet and

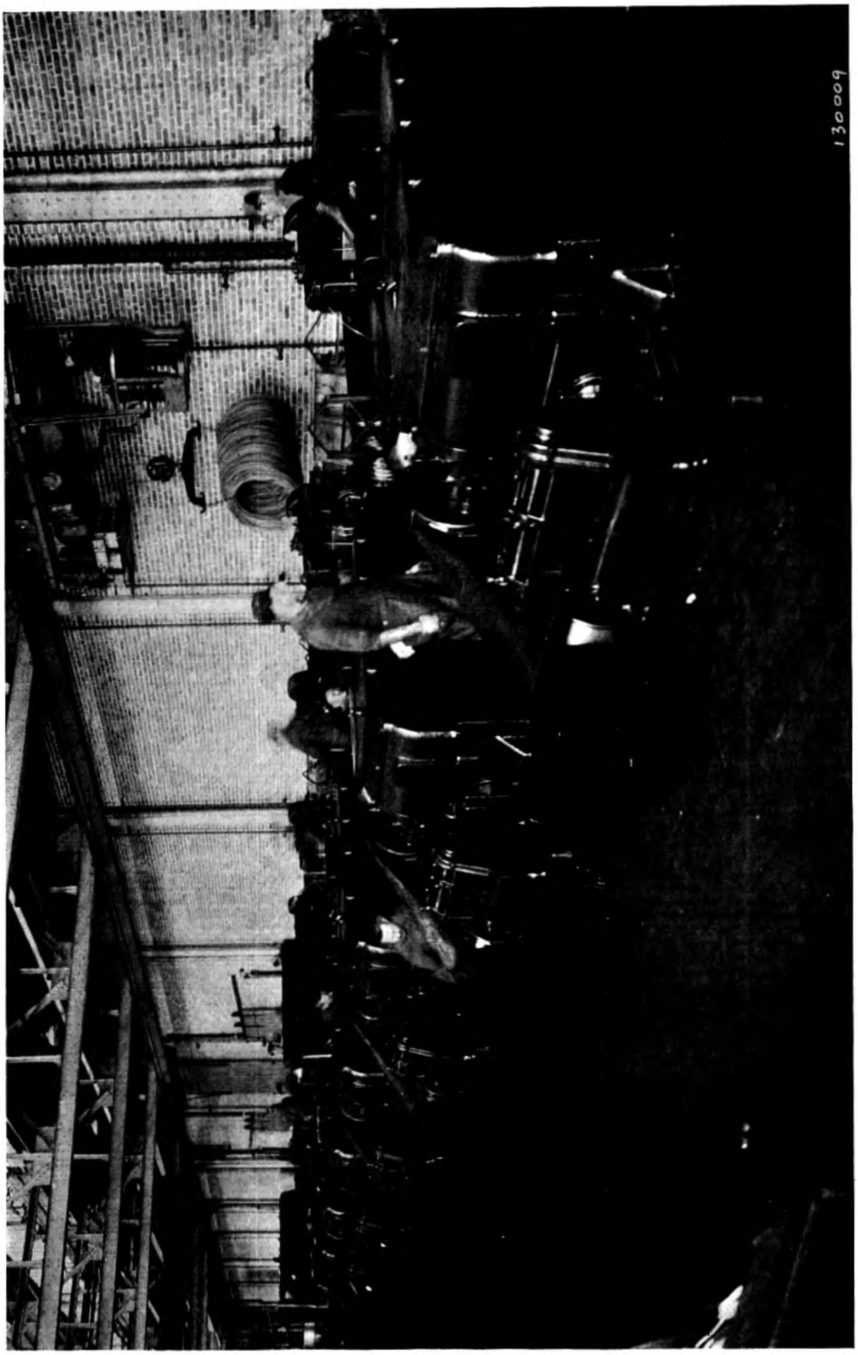


Fig. 11—Battery of No. 1 wire drawing machines

runs at 1500 to 2000 feet per minute as compared with 470 square feet for a commercial nine die machine running about 1000 feet per minute.

A battery of these machines costs much less than an installation of commercial machines of the same capacity, and in addition effects a considerable economy in floor space.

Fig. 12—Close-up of No. 1 machine

The commercial types of ten die intermediate machines for drawing cable wire require about 130 square feet of floor space as compared to 90 square feet for a twelve die multiple head machine. The former is a single unit machine and the latter a four unit machine operating at twice the speed and capable of producing about five times the output of the commercial equipment. This new multiple unit machine, Fig. 12A, costs more than regular equipment, but considering the four units, the cost is materially less per unit, and very much less on an output basis.

The magnet wire drawing machine is a high speed twelve die multiple head machine of eight wire drawing units occupying 90 square feet of floor space including the operating area. A close-up view of two heads of this machine is shown by Fig. 13. Fifty-one square feet of

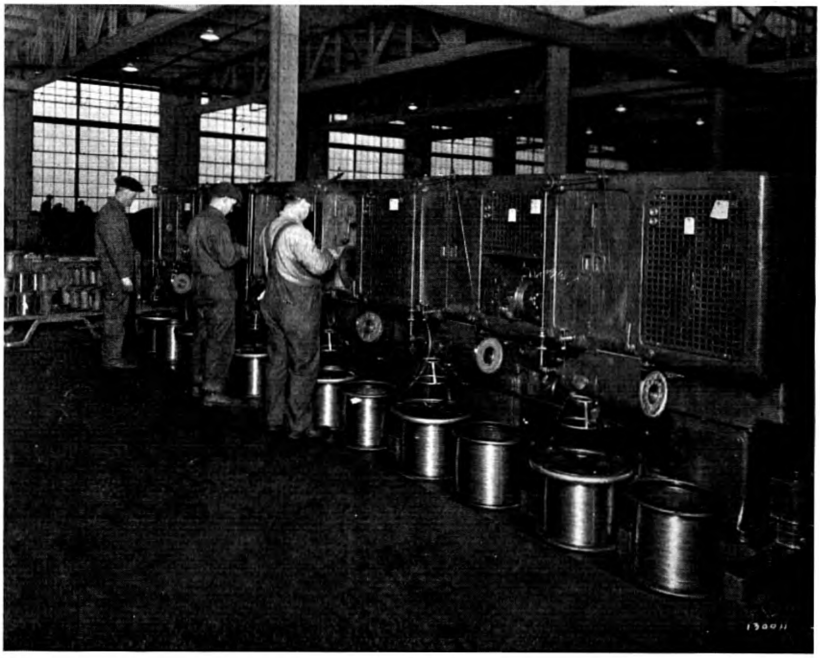


FIG. 12A.

floor space are required for a single unit (one head) commercial machine of the same die capacity. The saving in investment in this case is even greater than for the heavy and intermediate types of machines. The use of these compact machines and overhead monorail equipment for transporting material instead of using trucks with large aisles has permitted the installation of the wire drawing mill in less than one fourth of the building area which would have been required if commercial equipment had been purchased.

GENERAL PLANT FEATURES

The present connected load of the motors in the Rod and Wire Mill is about 6000 horse power for which it was necessary to enlarge the main power plant. A 700 foot tunnel connects the power plant with the Rod and Wire Mill in which are laid pipes for carrying hot and cold water, steam, gas, and air and lead covered power cables.

The basement under the Rod Mill houses the electrolytic equipment, control boards for the roughing and intermediate mills, pumps for cooling water, and exhaust fans connected with an air washer for removing the fumes from the Rod Mill. A tunnel which passes beneath the intermediate and finishing mills connects with a room which houses the drives for the four rod coilers, the coiler control boards, the finishing mill control board, and the main power panel. In the wire mill basement are six large tanks which hold the compound used to lubricate and cool the wire drawing dies. This compound is supplied under pressure to the wire drawing machines on the floor above and returns by gravity.

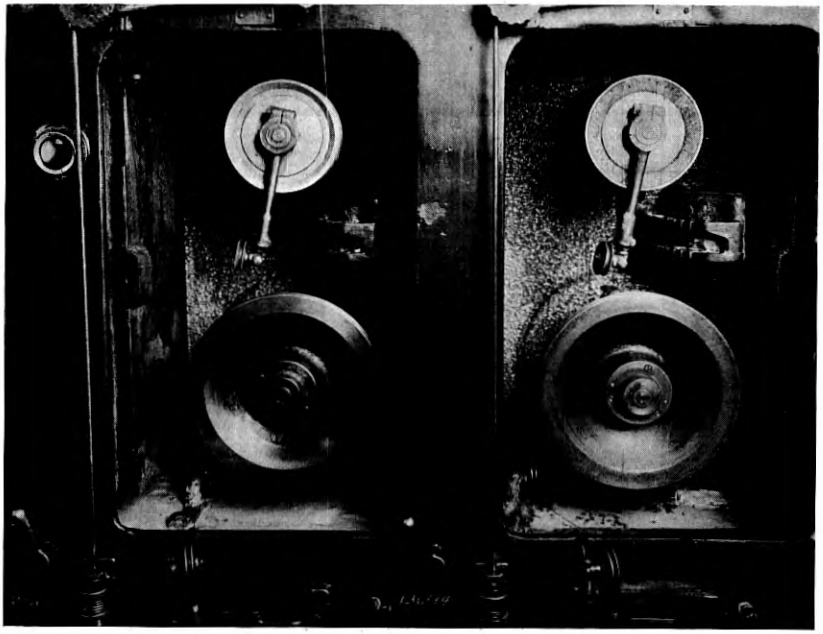


Fig. 13—Close-up view of units of No. 3 wire drawing machine

All the wire drawing machines are controlled by push buttons mounted on the machines, which connect with compensators in the basement. The 100 horse power motors driving the large wire drawing machines are mounted in a tunnel and are connected to the machines above by chain drive.

This arrangement permits accessibility for maintenance of the electrical equipment with a minimum of interference to production, prevents the wire drawing operators from having access to the electrical equipment, and reduces accident hazard to a minimum.

Developments in Wire Drawing Equipment and Methods

The Rod and Wire Mill just described was designed following a comprehensive survey of wire drawing processes and equipment used in this country and abroad. In connection with these studies, extensive laboratory investigations were undertaken relative to the characteristics of different types of commercial machines especially from

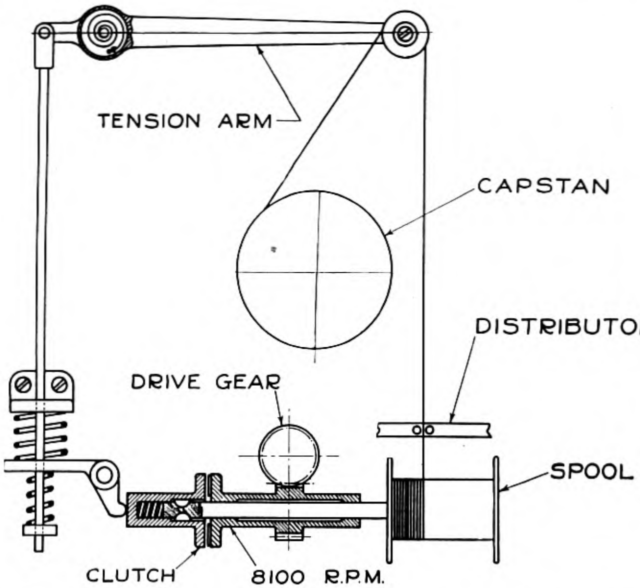


Fig. 14—Automatic tension mechanism—No. 3 wire drawing machine

the standpoint of operating efficiency, investment, and floor space requirements. As a result of these investigations, it developed that marked improvements could be effected if wire could be produced commercially at higher machine speeds and with more compact machine equipment.

While the design of the drawing mechanism in the new machine was very important, it was also deemed essential that the finished wire be taken up on spools instead of coils. After considerable experimental work, a sensitive take-up device was developed to permit spooling at a constant drawing speed.

This spooling mechanism is illustrated by Fig. 14 in which the spool spindle is driven by a slipping clutch member controlled through a tension arm, on which an idler pulley is located over which the wire passes on its way from the drawing capstan to the take-up spool. The take-up mechanism rotates the core of an empty spool at a speed synchronous with the speed of the wire as it leaves the drawing capstan. As the spool fills and the speed tends to increase, the wire on the tension arm tightens and compresses the tension arm against a spring adjusted for the proper gauge of wire. This in turn reduces the pressure of the clutch driving the take-up spindle, permitting the spool of wire to readjust its speed.

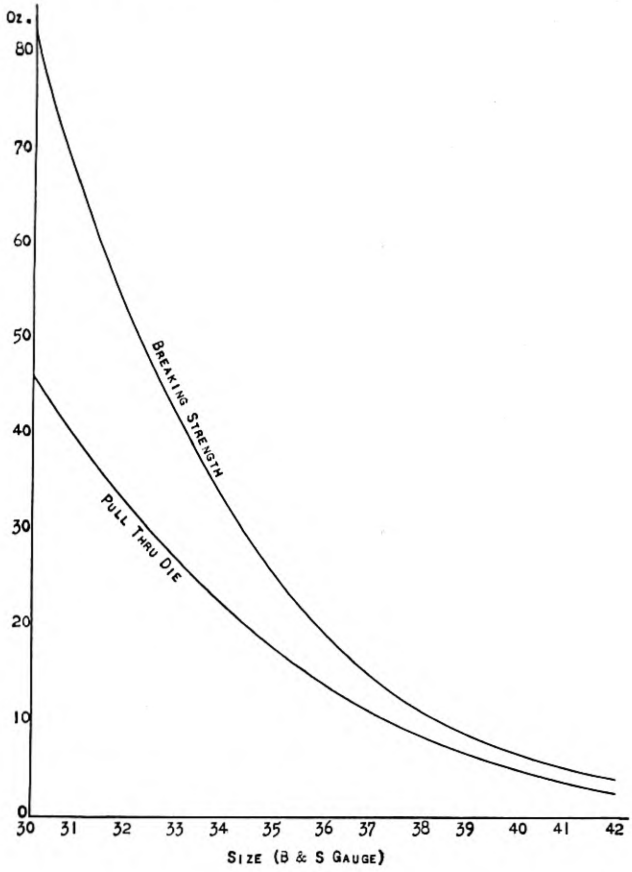


Fig. 15

This device is extremely sensitive as illustrated in the drawing of No. 42 B. & S. wire at 2000 feet per minute, in which case the control arm must be adjusted to operate between 90 and 150 grams, since the pull required is 87 grams and the breaking strength of the wire is 170 grams. This device is so flexible that it can be adjusted to a drawing tension of from 9 pounds for No. 25 wire to 3 ounces for No. 42 wire. Fig. 15 illustrates its operating range on wire sizes No. 30 to No. 42, showing the gradual narrowing of the limits as the sizes decrease. A larger machine used for drawing loop cable wire from No. 18 to No. 30 B. & S. gauges contains a similar mechanism.

The use of this sensitive device and a clutch which would slip without overheating as the spool filled, together with improvements in the wire drawing compound and the shape and quality of the diamond dies later described, permitted the drawing of wire at speeds ranging from 2000 to 3000 feet per minute.

Wire Drawing Compound

At low speeds it was discovered that the compound for lubricating wire drawing dies required little attention but as the speeds were increased the necessity for close analytical control was evident. The compound consists of an emulsion of soap, tallow, and water, the percentage of the soap and tallow being varied depending upon the size of wire and type of machine on which it is used.

It is important that the degree of emulsification [4] be carried far enough to break the tallow into particles about one micron in diameter, so that the material will stay in suspension in the water. If the tallow content is increased beyond a certain point, it holds in suspension in the solution a large amount of the copper dust which flakes off in a very fine state during the wire drawing operation and this clogs the dies and causes breakage during the wire drawing. Ordinarily this copper dust settles out of the solution while in the large cooling tanks and a considerable amount is salvaged in this manner.

Effect of Drawing on Copper

Tests were made to determine if the drawing of the smaller cable and all magnet wire sizes [5] in Brown and Sharpe (A.W.G.) steps was yielding the maximum reduction possible per die. These tests showed it was feasible to make much heavier than A.W.G. reductions at the first draft when annealed wire or soft copper rod was being drawn. It also showed that the elongation [6] of the rod or annealed wire was rapidly reduced to the drawing minimum after the first pass, and remained at that point throughout the process.

[4] "The Theory of Emulsions and Emulsifications," W. Clayton.

[5] A.W.G. ("American Wire" or "Brown and Sharpe" gauge) reductions are not used in converting the rod to line wire; these are generally specified in B.W.G. and N.B.S. gauges.

[6] See Figs. 16, 17, 18, and 19 showing the elongation of the rod or wire dropping to about 1½ per cent at the first die reduction and remaining practically constant.

Figure 16 illustrates the effect of a five die reduction on elongation and tensile strength. It may be seen that the elongation drops very rapidly at the first die when a reduction in area of about 42½ per cent is made, and the tensile strength increases rapidly because of the cold working of the metal.

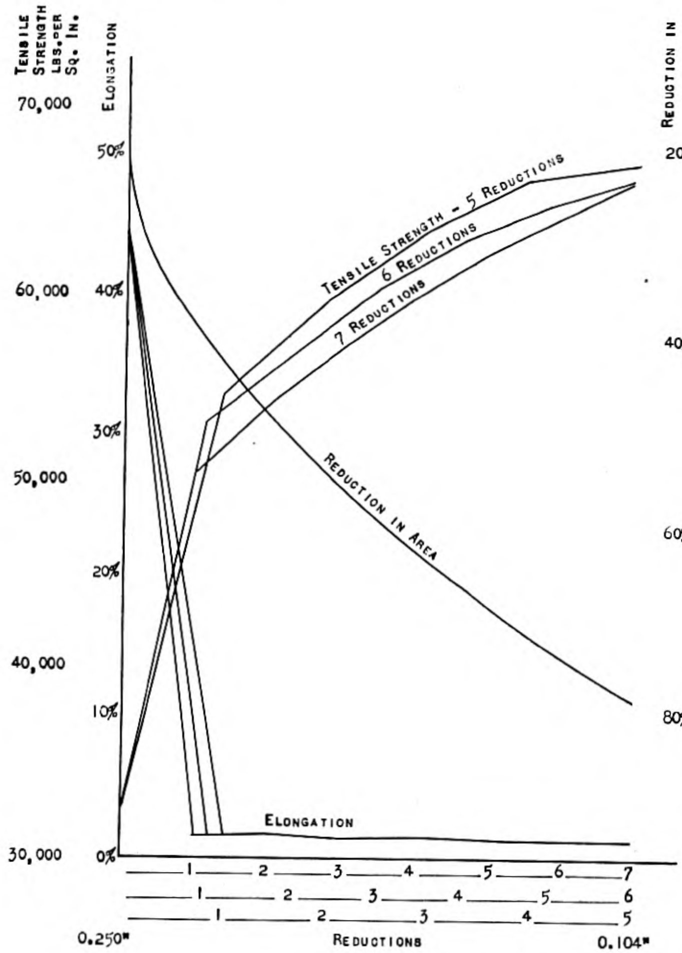


Fig. 16

This same figure shows the tensile strengths obtained when five, six, and seven die reductions are used to produce line wire of .104 diameter from the same supply. Here the elongation loss is about the same in each case, but the tensile strength is greater with the heavier reductions. The five die arrangement is satisfactory according to the results shown on the curve, but the heavy reduction at the first die often results in rough or slivered wire. The six die arrangement, therefore, gives the greatest factor of safety. The seven die arrangement is less satisfactory since the elongation and tensile strength in the finished wire are so close to the requirements.

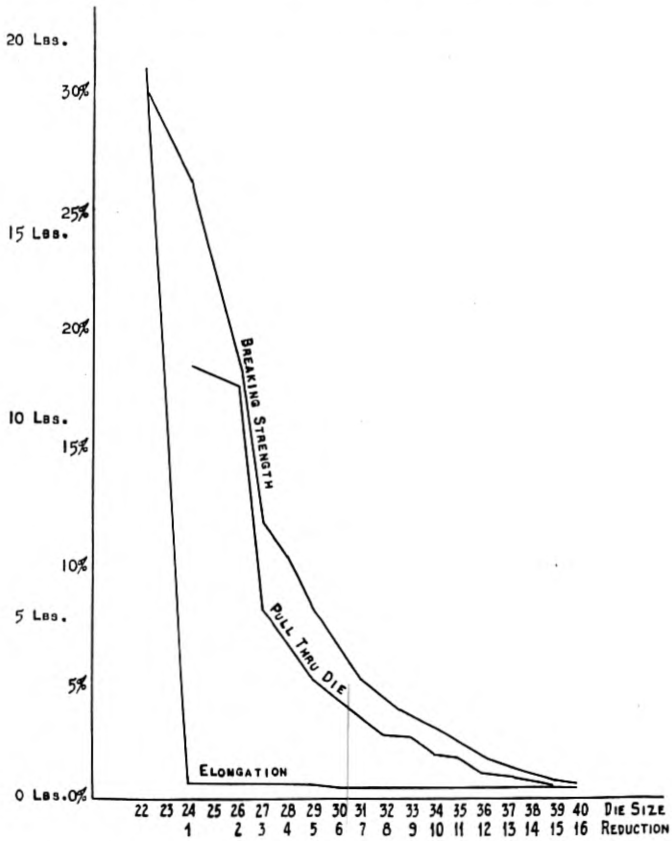


Fig. 17

The use of A.W.G. reductions for the finer sizes of cable and magnet wire provides flexibility since a change in the size of wire can be accomplished simply by increasing or reducing the number of dies used. Tests were conducted to determine the gain by using heavier reductions and annealing the wire before redrawing, and Fig. 17 shows the increased reduction possible at the first die when the metal is plastic. In this test, an annealed No. 22 gauge wire of 31 per cent elongation was reduced to No. 24, two gauges, in one draw. The soft copper permitted a double reduction at the first die, but the elongation dropped during the operation to less than 1 per cent; the second reduction on this test was from No. 24 to No. 26 gauge and the pull required for this pass practically coincides with the breaking strength of the wire. Wire drawing under such conditions is impractical because the annealing operation is much more expensive than drawing hard wire from No. 22 to No. 24 in two passes.

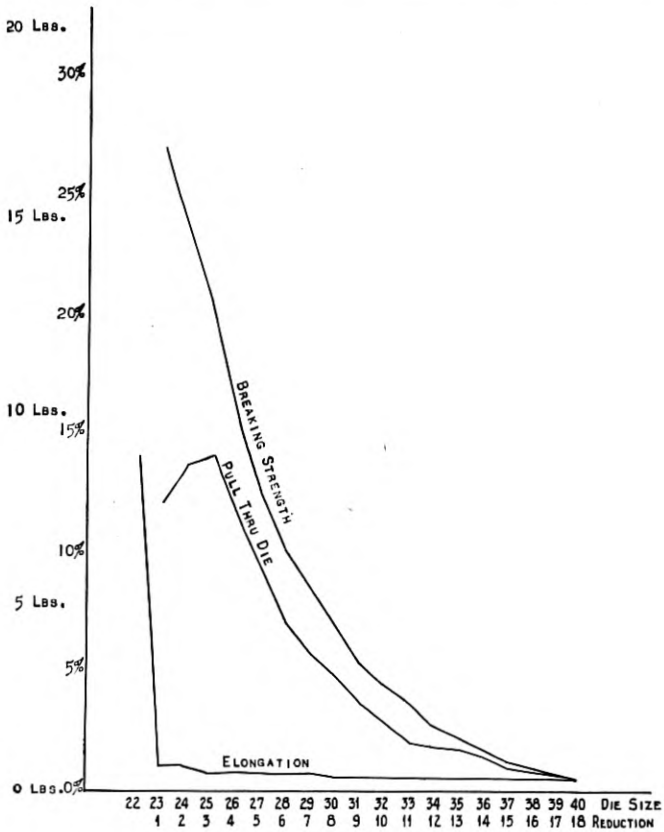


Fig. 18

Figure 18 illustrates the results obtained when drawing annealed wire with A.W.G. reductions. The large margin of safety between the pull required and the breaking strength of the material again disappears after two reductions. Fig. 19 [7] illustrates practical drawing

[7] Slight irregularities in the curves are due to variations from the mean in the diameters of the dies used during the test.

conditions adopted for drawing wire to finished sizes without annealing during the process.

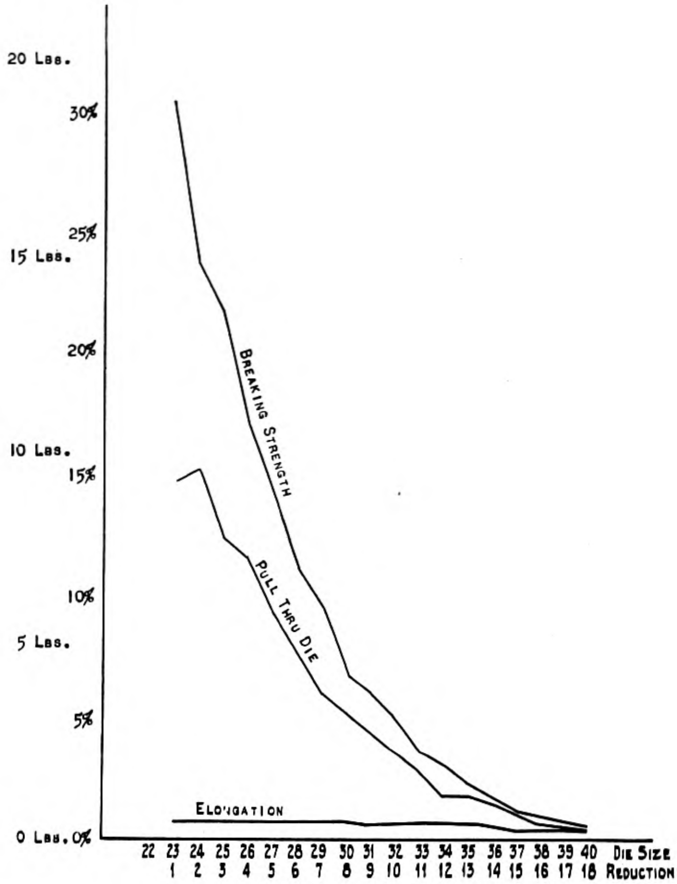


Fig. 19

Chilled Iron Dies

The dies used for drawing heavy wire are cast with a tapered hole from chilled cast iron and reamed to the desired size. When the die wears too large for a particular size of wire, it is reamed to a larger size and used in that manner until the die goes above the maximum size used. These dies, shown in Fig. 20, are used for drawing line and heavy gauge wire for which the cost of diamond dies would be excessive. Many alloy steel dies have been tested as substitutes for chilled iron dies for copper wire drawing, but so far have failed to replace them, due to excessive cost. For the wire sizes smaller than No. 16 down to as fine as No. 42 B. & S., diamond dies as described below are used.

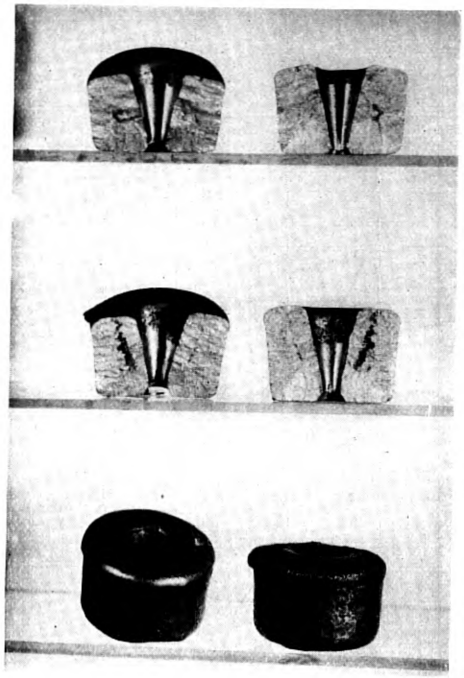


Fig. 20

Diamond Die Study

It was necessary to make an extensive study of the manufacture of diamond dies because dies through which wire could be satisfactorily drawn at low speeds failed to draw to gauge and without excessive breakage of the wire as the speeds were increased. At this time practically all commercial diamond drilling was done in Europe, Belgium being the hub of the diamond cutting industry, and the art was new to this country. The diamonds generally used for wire drawing dies are obtained from South Africa,[8] Australia, and Brazil, and made into diamond dies in Europe.

[8] The South African and Australian diamonds are the more suitable for wire drawing. There are two types of the former, the smooth brown premier which is not suitable for dies because of its tendency to crack and split, the other commonly known as the Jager, a product of the Jagerfontein mines. These stones, very irregular in contour and light gray to black in color, are most suitable for dies. The Australian diamonds are gray to brown to almost black in color and can be distinguished from the Jager. Many of the Brazilian diamonds are a dark gray similar to graphite in color and not being translucent are difficult to inspect for seams, cracks or inclusions.

In view of the difficulty of obtaining dies for drawing wire at high speeds and the large investment in dies required for the proposed wire mill, it was decided to undertake a laboratory investigation of the

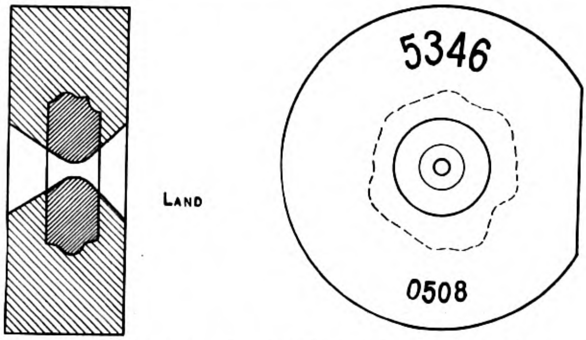


Fig. 21—Diamond wire drawing die (outline sketch)

manufacture of diamond dies suitable for drawing cable and magnet wire.

It was found that the dies suitable for high speed wire drawing required a differently shaped approach, a better polish, and a shorter land [9] than used for low speed drawing. In addition, the origin of the

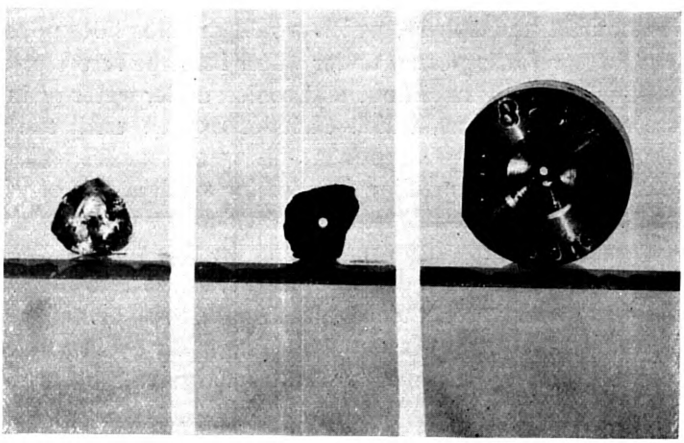


Fig. 22

stone, the shape of the diamond, and its setting are all very important because of the internal strain to which the die is subjected during the drawing operation.

[9] See Fig. 21.

It has not been possible to definitely establish any quantitative relationship as to the effect of high speed drawing on the wear of dies except that about the same number of million feet of wire may be expected from a properly lubricated die irrespective of the drawing speed. Under such conditions, the high speed die naturally runs a shorter time, but length of life is not the important factor; tonnage of a satisfactory quality with a minimum plant and labor investment is the prime consideration.

Figure 22 shows a diamond before drilling, a stone drilled and lapped, ready for mounting, and a die in the final mounting ready for use.

Figure 21 gives an outline of the shape of the working surfaces of a wire drawing die.

Annealing

Hard copper wire is obtained by using the wire as it comes from the wire drawing machine. This same wire may be softened by annealing, or medium-hard wire can be produced by annealing hard wire at such a point in the drawing operations that the final draws will give the desired degree of hardness.[10]

In a recent commercial type of annealing furnace, Fig. 23, wire may be bright annealed, but it requires a drying operation in order to remove the water through which it passes in leaving the furnace. The retorts of these furnaces are water-sealed and filled with steam to exclude the outside atmosphere, which would discolor hot copper. To obtain bright wire, it is passed under water into the retort to exclude the air and is generally taken out and cooled under water or in an atmosphere of steam or gas, which excludes oxygen until the wire is relatively cool.

A special steam seal annealing furnace for small spools of wire was developed on an experimental basis from which the wire was obtained bright annealed and free from moisture. In this furnace the spools were submerged in water to displace the air, raised into the charging end which was under water, thence to the muffle to be heated, and then along a cooling tube to the discharge opening. Air was excluded from the retort and cooling chamber at the discharge end by means of a steam jet.

The success of the small furnace led to the construction of a larger machine (Fig. 24) for annealing cable wire on spools. The spools are placed in perforated metal baskets which are charged into the furnace at a specified time interval, pushing each other through the retort and along the cooling tube to the discharge end.

[10] "Experiments in the Working and Annealing of Copper," F. Johnson, *Journal Institute of Metals*, Volume XXVI, No. 2, 1921.

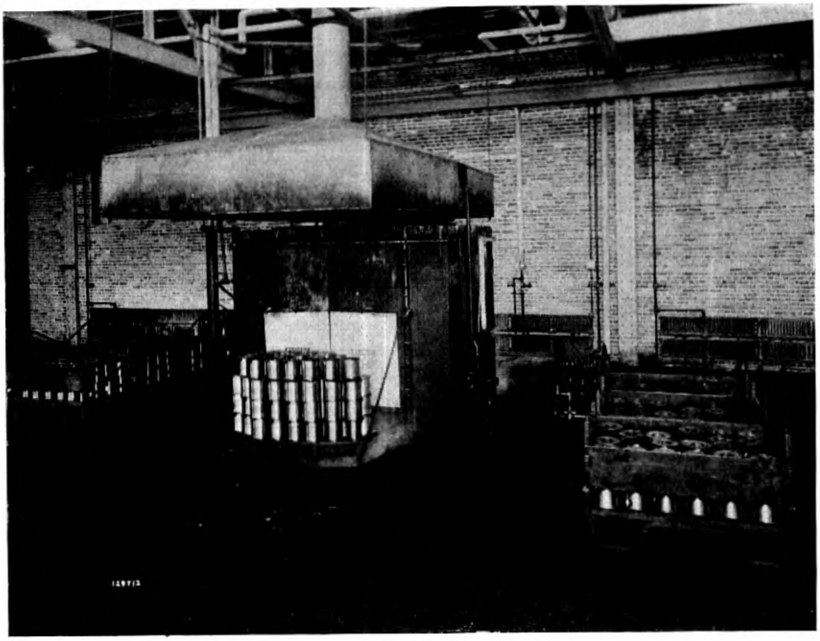


Fig. 23—Water seal annealing furnace

Inspection of Raw Material and Finished Product

Wire bar made from electrolytic refined copper is used as a material in the manufacture of wire. This material is practically free from silver and other elements which ordinarily exist in the ore, and which have a detrimental effect on the electrical or physical properties of the finished product. A small percentage of silver [11] seriously affects the annealing qualities of the wire. Traces of other impurities have a very detrimental effect on the wire drawing properties. During the refining process, the molten bath is oxidized in order to carry off the foreign material in the form of slag, and it is very important that the oxygen content be later reduced to a very small point if bars of proper set are desired. Fig. 25 shows three photomicrographs of wire bar containing varying amounts of cuprous oxide.[12] Ordinarily the surface condition on top of the bar is a good index of the oxygen content. If the bar is level set or slightly convex on top, it is usually a satisfactory material. If it is low set or concave, it usually contains a large amount of copper

[11] "Effects of Silver on the Recrystallization Temperature of Copper," Caesar and Gerner, *A. S. M. E.*, Volume 38, 1916.

[12] "Microscopic Structure of Copper," H. P. Pulsifer, *Mining and Metallurgy*, January, 1926.

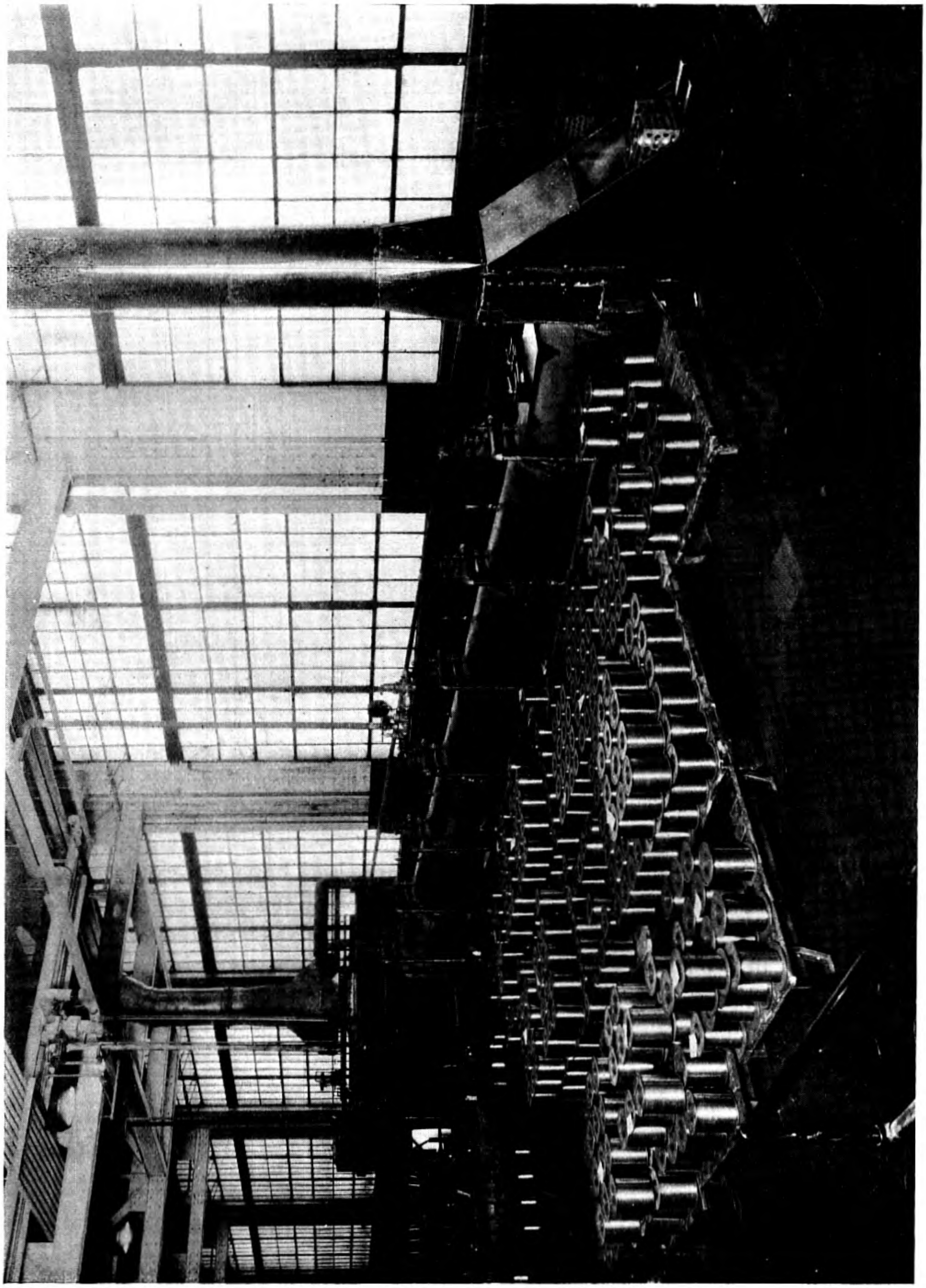
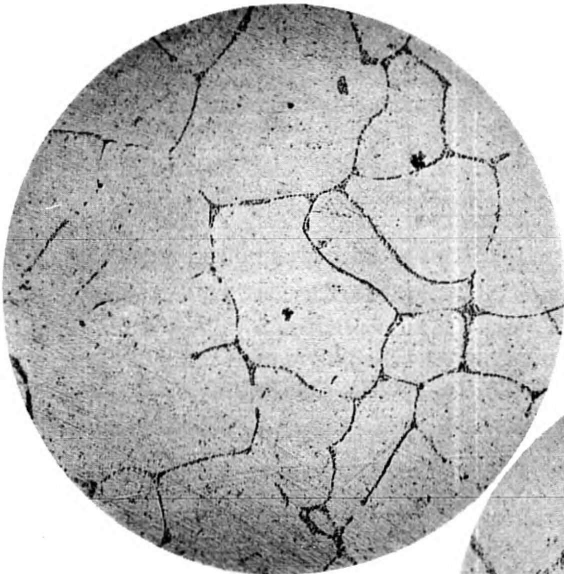
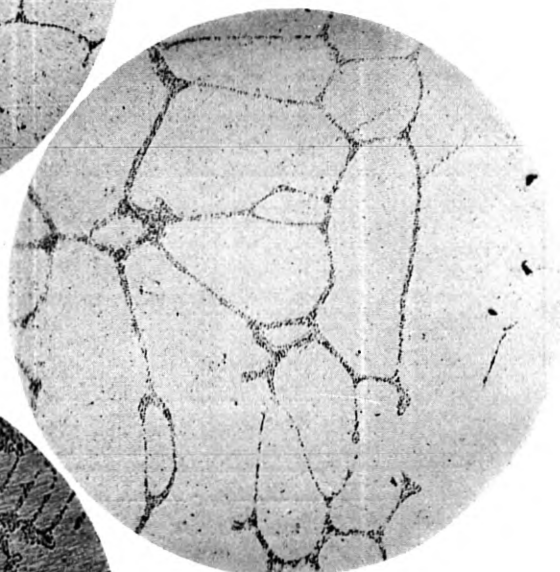


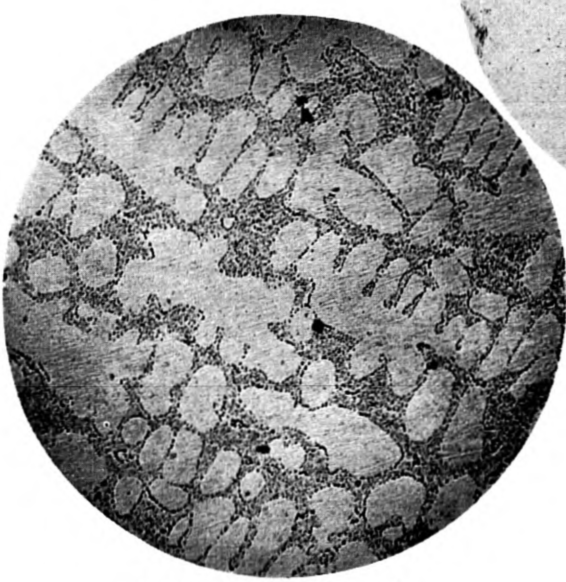
Fig. 24—Steam seal annealing furnace



A. High Set—Oxygen .035%



B. Level Set—Oxygen .050%



C. Low Set—Oxygen .12%

Fig. 25—Photomicrographs of wire bar (magnification × 100)

oxide, which caused the metal to shrink in solidifying.[13] When excessive shrinkage occurs it has an adverse effect during the rolling operation.

The finished wire is inspected for dimensional limits, tensile strength, elongation, and surface condition. The limits for 42 B. & S. gauge wire (.002475 in.) are .00245 in. minimum and .0025 in. maximum.

Conclusion

The establishment of this industry as a part of the plant at Chicago represents the combined effort of a large number of inventors, engineers, designers, and mechanics. While the actual plant was built within a comparatively short period, the advances which have been made in the art represent several years' effort. Briefly, the development of compact and high speed wire drawing machines has required a much smaller investment in buildings and equipment as compared with a plant of the same capacity using commercial equipment. A comparison of the relative floor area, based upon the conventional and the improved types of wire drawing equipment, is illustrated by Fig. 26. The supervisory force in charge of the operation of this new mill must be given a considerable share of the credit for its successful operation.

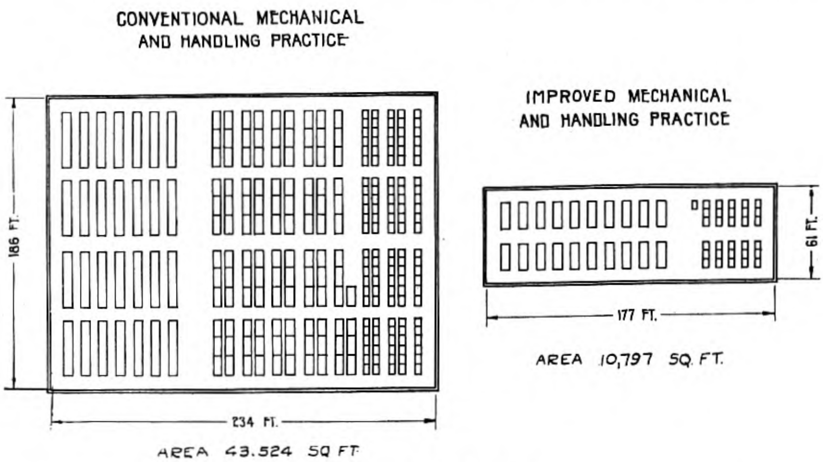


Fig. 26—Wire drawing plant

[13] "Copper Refining," Lawrence Addicks. "Metallurgy of Copper," H. O. Hofman.

An Analyzer for the Voice Frequency Range

By C. R. MOORE and A. S. CURTIS

[EDITORIAL NOTE: The frequency analyzers described in this paper and in the paper immediately following, demonstrate in an unusual manner how a single fundamental principle may be employed to accomplish quite dissimilar results. The analyzers described in both papers employ a resonating element of fixed frequency and translate the wave components under study to this frequency by heterodyning them with the output of variable frequency oscillators. In the analyzer described in the first paper, the wave components under study are translated to a higher frequency while in that described in the second paper the translation is downward to a lower frequency. In view of these differences in design it is desirable to call particular attention to the reasons which have led to the working out of the two designs.

The analyzer discussed by Moore and Curtis has been so designed as to sweep through the voice frequency range to as high as 5,000 cycles by the manipulation of a single control. To accomplish this end, it was found desirable to heterodyne upward by employing a variable frequency oscillator of considerably higher frequency than 5,000 cycles. The frequency of this oscillator can be varied continuously throughout the range from about 11,000 cycles to 16,000 cycles, and the fixed frequency resonating element is tuned to about 11,000 cycles. As translation of the wave under study to a higher frequency range reduces the percentage separation of the various components, it was necessary to choose a very sharply tuned resonating element. This takes the form of a steel rod which is loosely coupled magnetically to a driving circuit at one end and a registering circuit at the other. As the modulator used to accomplish the heterodyning process produces many frequencies other than the first of the "sum" and "difference" terms, it has been necessary to choose the frequency ranges such that all undesired frequencies which can not be made extremely small will be well removed from the single difference frequency under observation.

The analyzer described in the paper by Landeen is capable of working over the range from about 3,000 cycles to 100,000 cycles. A requirement of this design was that very high resolution be obtained. To assist in accomplishing this end, the frequencies under study are translated downward in the frequency scale to the resonating element which consists of a circuit tuned to 800 cycles. This downward translation increases the percentage difference of frequency separation of the components under study. Because of the great range of frequencies covered by the analyzer it is not possible to have sum and difference terms other than those of the second order fall outside of the range of sensitivity of the resonator. The modulator has therefore been so designed as to preclude formation in the higher order terms. To increase its discrimination, the analyzer makes use of two tuned circuits and amplifiers arranged in tandem and placed before the modulator. The frequency to which these circuits are tuned must of course be variable and is set to coincide with the component under study.]

THE present analyzer was designed to aid in the solution of certain problems arising in the study and development of commercial telephone transmitters. These problems require high discrimination and the accurate measurement of frequency components in the presence of much larger components.

The present analyzer differs fundamentally from that described by R. L. Wegel and one of the present authors about two years ago.[1]

[1] "An Electrical Frequency Analyzer," by C. R. Moore and R. L. Wegel, *A. I. E. E. Journal*, September 1924; *Bell System Technical Journal*, October 1924.

It is of the type employing a single resonating element of fixed frequency, the component waves under study being translated in frequency to it by heterodyning with the output of a variable frequency oscillator. It was deemed essential that the analyzer be so designed as to cover the entire voice range up to 5,000 cycles by the manipulation of a single control. In the presenta analyzer this control is the variable condenser of the heterodyning oscillator. Two of these analyzers have been built and are in use in the Bell Telephone Laboratories.

The development of this new form of analyzing device was undertaken only after a careful review of existing types. Such a study led to the conclusion that none of the available forms was applicable to the solution of our type of problem. This problem which had to do

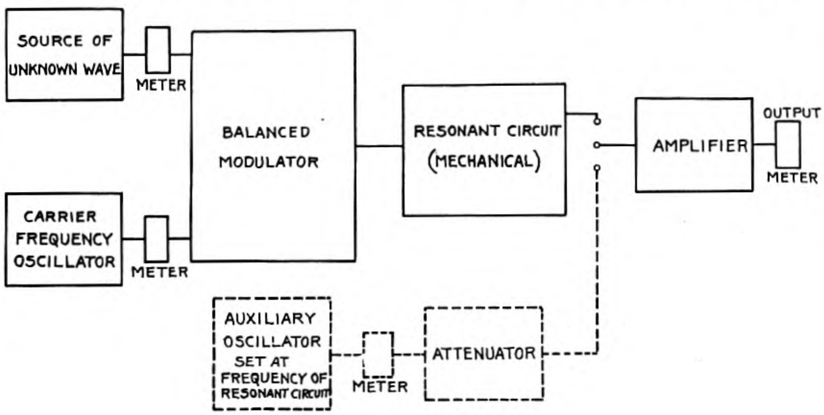


Fig. 1—Functional Diagram

with the study of commercial telephone transmitters did not require the analysis of a complex wave containing a fundamental frequency and its associated harmonics. What we were interested in was the measurement of the transmitter output at a particular frequency in the presence of a much greater output at other frequencies. For example, the measured component may be only one per cent of the magnitude of the disturbing component and separated from it by only 30 to 40 c.p.s. It is evident that if the frequency of the component which we wish to measure is close to that of one of the disturbing components and much smaller in magnitude a very high degree of frequency discrimination on the part of the analyzer is necessary.

A functional diagram of the analyzer is shown in Fig. 1. The wave to be analyzed is impressed at the "voice" input terminals of a balanced modulator. The variable carrier frequency is also supplied to the modulator and in conjunction with the wave to be analyzed produces the familiar upper and lower sideband frequencies in the modulator output.[2] The modulator output is connected to a resonant element, the natural frequency of which is left unchanged during the analysis, and the response of the resonant element is measured by a suitable amplifier and a meter. The tuned element of the analyzer makes use of longitudinal vibrations in a steel bar which in the apparatus herein described has a natural frequency of approximately 11,000 c.p.s. The frequency range of the carrier oscillator is from the natural frequency of the resonant element to 5,000 c.p.s. above. It will readily be seen that if, for instance, there is a 1,000 c.p.s component in the unknown wave and the carrier frequency oscillator is set at a frequency of 1,000 c.p.s. above the natural frequency of the resonant element, the lower sideband or difference component from the modulator will be at the frequency of this resonant element. The process of analysis is then to vary the frequency of the carrier oscillator gradually and to determine the output from the resonant element at each desired frequency. Inasmuch as the frequency range of the carrier oscillator is relatively small, the entire variation of frequency can be accomplished by a single air condenser. Therefore, the frequency setting of the analyzer may be varied continuously instead of in discrete steps. The frequency calibration chart of the carrier oscillator is arranged to show the input frequencies of the unknown wave to which a given frequency of this oscillator will correspond rather than to indicate the frequency of the oscillator itself.

It has been found convenient, for comparing the magnitudes of the various frequency components in the unknown wave, to use an auxiliary oscillator supplying current to a potential attenuator. The frequency of this oscillator is maintained at the frequency of the tuned element. The input terminals of the amplifier may be connected either to the output of the resonant circuit or to the output of the potential attenuator. The procedure is to note the deflection on the output meter of the amplifier produced by the resonant element and then switch the amplifier to the attenuator and to reproduce this deflection. The frequency components can readily be compared in this manner and their relative magnitudes determined directly in *TU* by reference to the attenuator dial setting.

In order to give a clearer idea of the operation of the analyzer the pertinent theory of the vacuum tube modulator will be discussed in the Appendix.

[2] *A. I. E. E. Journal*, April 1921—"Carrier Current Telephony and Telegraphy," by E. H. Colpitts and O. B. Blackwell.

The modulator used employs two vacuum tubes and its circuit is arranged to suppress the carrier frequency together with certain higher order modulation components in the modulator output. These together with other higher order modulation components which are not eliminated in this type of balanced modulator could produce false indications of frequency components in the wave to be analyzed, but it will be shown in the Appendix that these errors may be reduced to any desired extent by keeping the magnitude of the wave to be analyzed as low as is consistent with securing satisfactory meter readings.

The suppression of the carrier wave is desirable in that it makes it possible to carry the analysis to lower frequencies than could be done if the carrier frequence were present. When analyzing low frequencies, the frequencies of the carrier wave and the lower sideband approach each other and if the relatively large carrier were present in the output of the modulator it would tend to obscure the results.

It will readily be seen that since the resonant frequency of the tuned circuit is 11,000 c.p.s. and since the frequency discrimination at low frequencies depends upon the sharpness of resonance of this circuit, extremely sharp tuning is necessary. The considerations here differ somewhat from the ordinary considerations in tuned circuits where the effect of the resonance depends upon a percentage departure from the resonant frequency. The reactance-resistance ratio, Q, which is in common use in the treatment of electrical circuits, gives a measure of what we may call the percentage sharpness of tuning of a circuit, that is, with a given value of Q, a given percentage departure from the resonant frequency will cause the same loss independent of the resonant frequency. While it is possible at frequencies from 10,000 to 20,000 c.p.s. to obtain higher values of Q than at frequencies from 100 to 1,000 c.p.s., it is not feasible to secure the same loss with a given departure in cycles from the resonant frequency at high frequencies as it is at low frequencies. It is evident that in this method of analysis we are not concerned with a percentage departure in frequency from the resonant frequency of the tuned circuit but are concerned with the loss in transmission through this circuit per cycle departure from the resonant frequency. Therefore, inasmuch as we would desire to have good discrimination between a frequency of 100 c.p.s. and one of 110 c.p.s., the requirements of the 11,000 c.p.s. resonant circuit are extremely rigid.

Some consideration was first given to the design of an electrical network which would give sufficiently sharp tuning. At best, such a network required a considerable number of coils and condensers and these coils would require higher values of Q than could be obtained economically. Moreover, inasmuch as this selective circuit would consist of a number of highly resonant elements, it would be rather questionable whether these elements would all be affected alike by ordinary variations in room temperature. Some experiments were then made using mechanical resonance and these have given a very satisfactory solution of the problem.

The resonant element now in use consists of a steel rod clamped at the center having the magnetic element of a telephone receiver at each end with its poles separated a few mils from the end of the rod as shown in Fig. 2. One of the receiver units is connected to the output of the modulator and is used for driving the bar while the other receiver unit

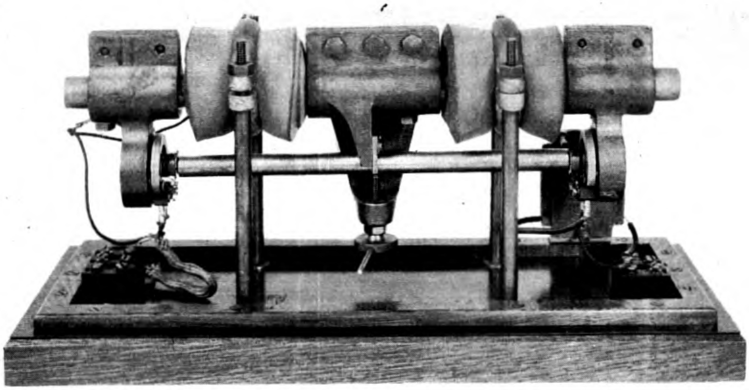


Fig. 2—Mechanical resonant circuit

is connected across the input of a suitable amplifier having a thermocouple and meter at its output. In order to minimize the effect of other extraneous frequency components the amplifier is tuned to have its maximum efficiency at the resonant frequency of the bar. The bar is approximately 9 inches long and resonates to longitudinal vibrations at 11,350 c.p.s. A frequency response curve of the bar, showing variation of the output of the driven receiver with frequency, is shown in Fig. 3. It will be seen that a departure of 10 c.p.s. from the resonant frequency gives a loss of over 25 *TU* corresponding to a voltage ratio of approximately one to twenty. Therefore, even when frequencies in the unknown wave are as low as 50–100 c.p.s. the frequency discrimination is quite satisfactory. It may be of interest to note that the value of the reactance-resistance ratio Q, calculated from the curve, is about 15,000 whereas the construction of an electrical inductance to operate at this frequency having a value of Q over 200 would be difficult and expensive.

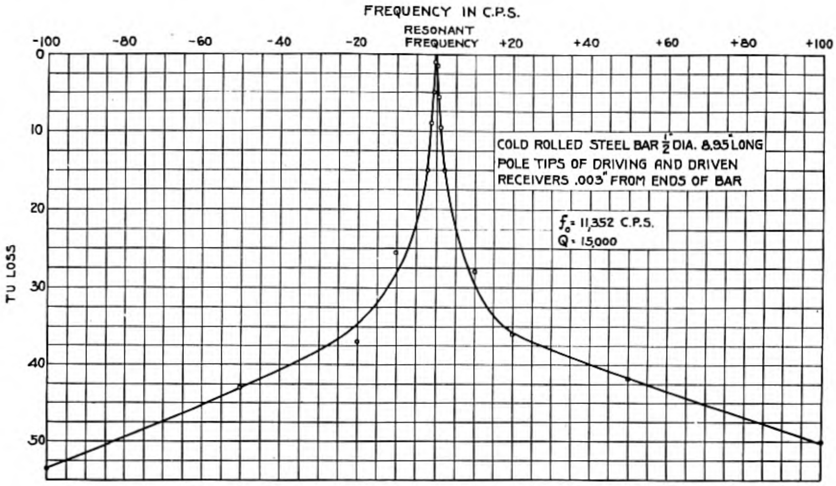


Fig. 3—Response-frequency characteristics of resonant bar

One of the two heterodyne frequency analyzers is built as a self-contained unit mounted on wheels and contains its own "A" and "B" batteries for the vacuum tubes. Fig. 4 shows the top view of the panel where the necessary controls and meters are situated and Fig. 5 shows the complete assembly with some of the compartments open. A schematic diagram of the complete circuit is shown in Fig. 6.

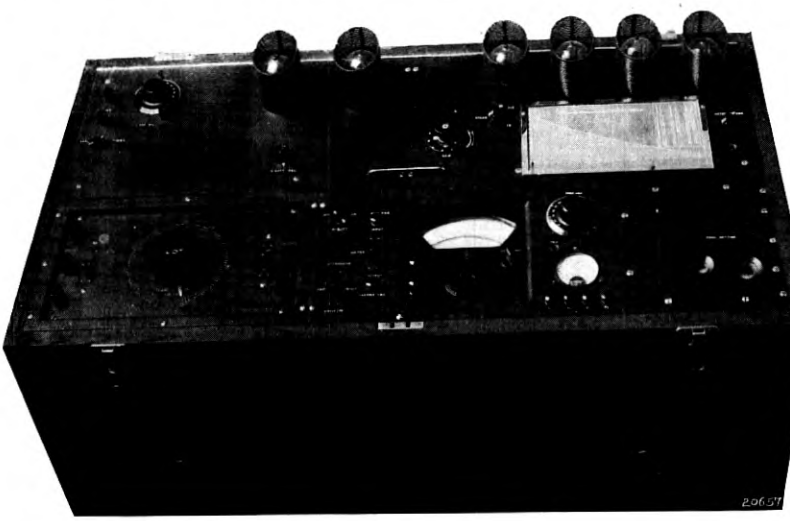


Fig. 4—Top view of panel

In addition to its use for determining the relative magnitudes of the components in an electrical wave, the analyzer is also useful in making acoustic measurements. It was primarily developed for the study of what we may term the frequency distortion of transmitters; it may also be used in the measurements of non linear distortion. As may be

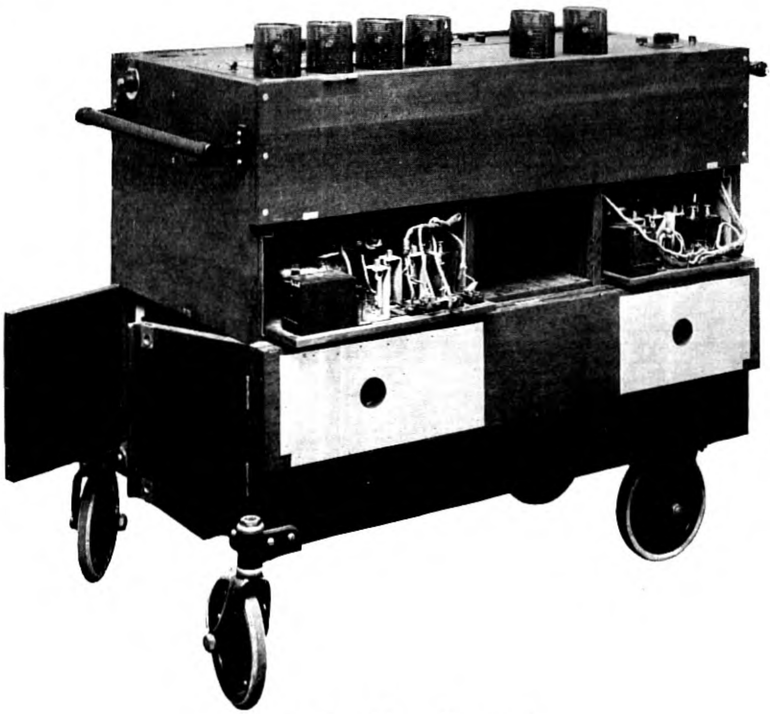


Fig. 5—Portable analyzer unit

inferred, frequency distortion is a departure of the wave form of the electrical output from that of the acoustic input due to the fact that the transmitter does not respond equally to acoustic forces of the same magnitude over the frequency range under consideration. Non-linear distortion is the distortion produced by the fact that the voltage across the transmitter at any particular frequency is not a linear function of the magnitude of the impressed acoustic force. This type of distortion usually manifests itself by the production of frequency components in the electrical output which are not present in the acoustic input. The analyzer may be used for quantitative determinations of this non-linear distortion and in such measurements high frequency discrimination and ability to measure frequency components of widely different magnitudes are very valuable.

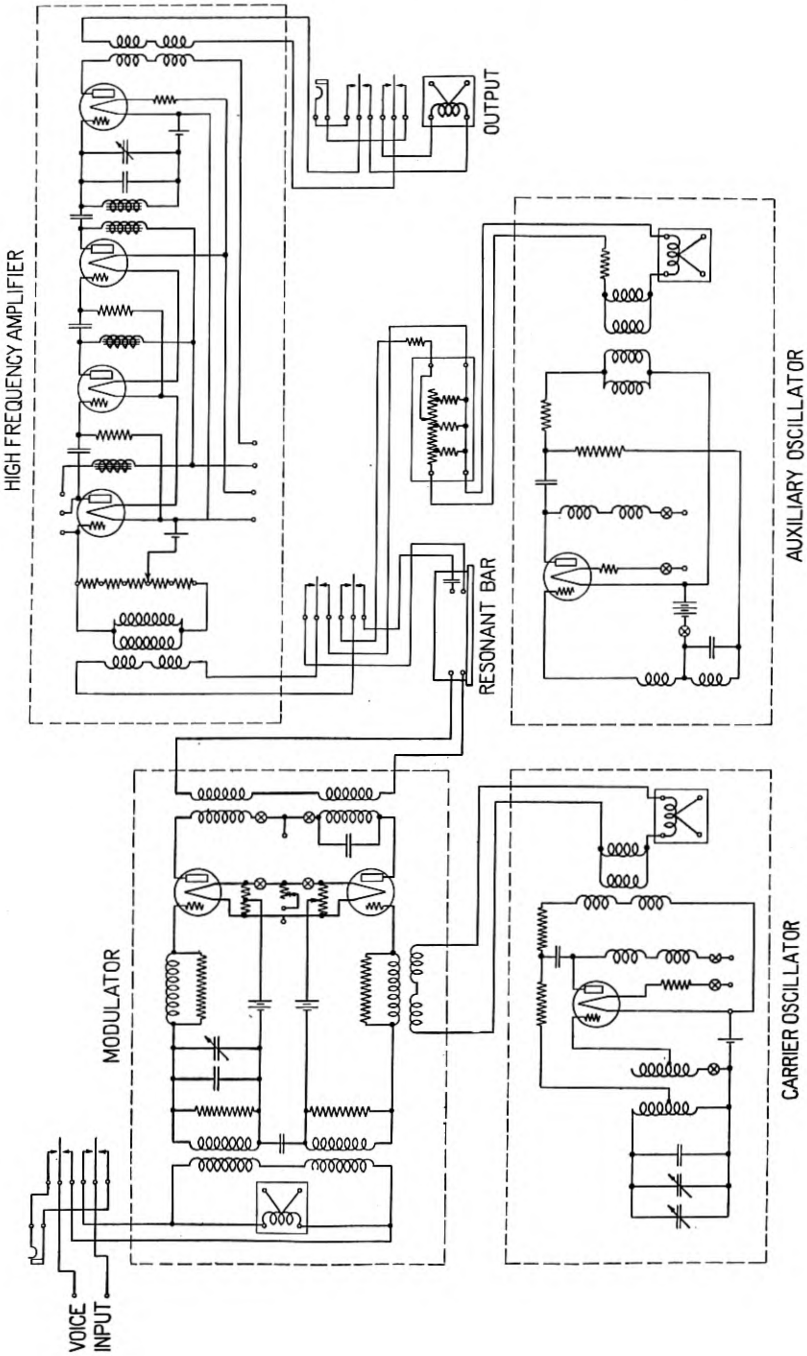


Fig. 6—Schematic diagram of analyzer circuit

A number of other uses of the analyzer in acoustic measurements might be cited; for instance, the output of a transmitter at one particular frequency may be measured in the presence of room noises or other disturbing sounds. In general its high selectivity, convenience of operation and portability make the analyzer an extremely valuable instrument in a wide variety of acoustic and electrical measurements.

Appendix

Inasmuch as the incorrect operation of the modulator could give false indications of frequency components in the wave to be analyzed, it may be of interest to take up some of the more important aspects of the theory of the vacuum tube plate current modulator as applied to this analyzer.[3] In general, the output of a vacuum tube is of the form

$$I_1 = A_0 + A_1E + A_2E^2 + A_3E^3 + A_4E^4 + \text{etc.,} \tag{1}$$

where E is the voltage applied between the cathode and grid of the tube and the coefficients A_0, A_1, A_2, A_3, etc., depend upon the average potential of the grid, the constants of the tube itself and the total impedance of the output circuit to the various frequency components appearing in the output. If we apply two voltages of the form $P \cos (pt - \theta)$ and $Q \cos (qt - \phi)$ simultaneously to the input of a vacuum tube, the above general expression will have the form

$$\begin{aligned} I_1 = A_0 &+ A_1[P \cos (pt - \theta) + Q \cos (qt - \phi)] \\ &+ A_2[P \cos (pt - \theta) + Q \cos (qt - \phi)]^2 \\ &+ A_3[P \cos (pt - \theta) + Q \cos (qt - \phi)]^3 \\ &+ A_4[P \cos (pt - \theta) + Q \cos (qt - \phi)]^4, \text{ etc.} \end{aligned} \tag{2}$$

For simplicity let $a = P \cos (pt - \theta)$ and $b = Q \cos (qt - \phi)$ and then expanding algebraically we obtain

$$\begin{aligned} I_1 = A_0 + A_1(a+b) + A_2(a^2+2ab+b^2) + A_3(a^3+3a^2b+3ab^2+b^3) \\ + A_4(a^4+4a^3b+6a^2b^2+4ab^3+b^4) + \text{ etc.} \end{aligned} \tag{3}$$

Now substituting $P \cos (pt - \theta)$ and $Q \cos (qt - \phi)$ for a and b respectively and simplifying trigonometrically, the frequencies appearing in the output of the tube are shown in Table 1.

If we employ a balanced modulator of the type shown in Fig. 7, certain frequency components shown in the table are eliminated. In this modulator the voice input transformer is so connected as to

[3] For a more complete discussion of the vacuum tube modulator see *Proceedings I. R. E.*, April 1919; "A Theoretical Study of the Three Element Vacuum Tube," by J. R. Carson.

TABLE 1

Terms in the First Four Orders of Modulation

$P \cos (pt - \theta)$ and $Q \cos (qt - \phi)$

General Expression for Current Output

$I_1 = A_0 + A_1E + A_2E^2 + A_3E^3 +$ etc.

Frequency $\omega = 2\pi f$	Coefficients			
	A_1	A_2	A_3	A_4
0		$1/2\ P^2$, $1/2\ Q^2$		$3/8\ P^4$, $3/2\ P^2Q^2$, $3/8\ Q^4$
$(pt - \theta)$	P		$3/4\ P^3$, $3/2\ PQ^2$	
$(Qt - \phi)$	Q		$3/4\ Q^3$, $3/2\ P^2Q$	
$(2pt - 2\theta)$		$1/2\ P^2$		$1/2\ P^4$, $3/2\ P^2Q^2$
$(2qt - 2\phi)$		$1/2\ Q^2$		$1/2\ Q^4$, $3/2\ P^2Q^2$
$(pt - \theta) \pm (qt - \phi)$		PQ		$3/2\ P^3Q$, $3/2\ PQ^3$
$(3pt - 3\theta)$			$1/4\ P^3$	
$(3qt - 3\phi)$			$1/4\ Q^3$	
$(2pt - 2\theta) \pm (qt - \phi)$			$3/4\ P^2Q$	
$(pt - \theta) \pm (2qt - 2\phi)$			$3/4\ PQ^2$	
$(4pt - 4\theta)$				$1/8\ P^4$
$(4qt - 4\phi)$				$1/8\ Q^4$
$(3pt - 3\theta) \pm (qt - \phi)$				$1/2\ P^3Q$
$(2pt - 2\theta) \pm (2qt - 2\phi)$				$3/4\ P^2Q^2$
$(pt - \theta) \pm (3qt - 3\phi)$				$1/2\ PQ^3$

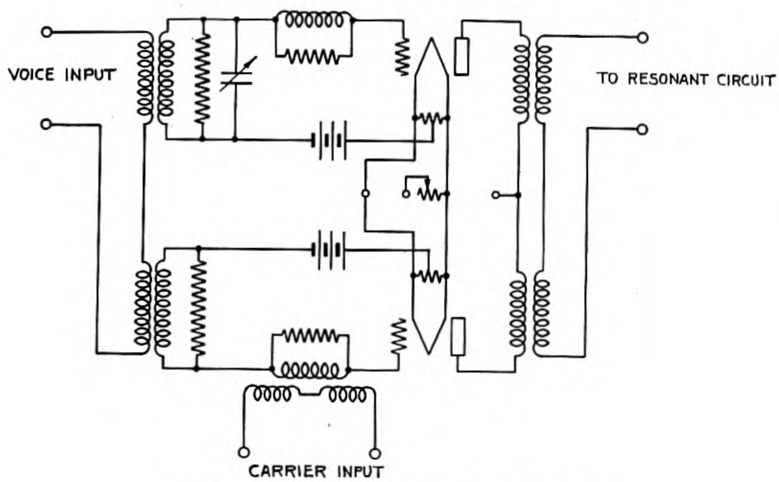


Fig. 7—Schematic diagram of modulator

impress instantaneous voltages of opposite signs on the two tubes, while the carrier windings impress instantaneous voltages of the same sign on the tubes. The output transformer is connected so that the algebraic difference of the two plate currents appears in the output of the modulator. Equation (3) shows the output of one of the tubes, but the output for the other tube in which the voice input is reversed will be

$$I_2 = A_0 + A_1(a-b) + A_2(a^2-2ab+b^2) + A_3(a^3-3a^2b+3ab^2-b^3) + A_4(a^4-4a^3b+6a^2b^2-4ab^3+b^4) + \text{etc.} \tag{4}$$

If the circuits of the two tubes are exactly balanced, then since the output of the modulator is so connected that only the algebraic difference of the two plate currents will appear, all of the frequencies arising from the terms of like sign in equations (3) and (4) will be suppressed or, subtracting equation (4) from equation (3), the resultant current will be

$$I_1 - I_2 = +2A_1b + A_2(+4ab) + A_3(+6a^2b+2b^3) + A_4(+8a^3b+8ab^3). \tag{5}$$

The frequency components arising from the various terms of equation (5) are shown in Table 2.

TABLE 2

FREQUENCY COMPONENTS ARISING FROM ALGEBRAIC DIFFERENCE OF $I_1 - I_2$

$2\ A_1Q \cos (qt - \phi)$
$2\ A_2PQ \cos [(pt - \theta) \pm (qt - \phi)]$
$3\ A_3P^2Q \cos (qt - \phi)$
$3/2\ A_3P^2Q \cos [(2pt - 2\theta) \pm (qt - \phi)]$
$1/2\ A_3Q^3 \cos [3qt - 3\phi]$
$3/2\ A_3Q^3 \cos (qt - \phi)$
$A_4P^3Q \cos [(3pt - 3\theta) \pm (qt - \phi)]$
$3\ A_4P^3Q \cos [(pt - \theta) \pm (qt - \phi)]$
$3\ A_4PQ^3 \cos [(pt - \theta) \pm (qt - \phi)]$
$A_4PQ^3 \cos [(pt - \theta) \pm (3qt - 3\phi)]$

In the analyzer the only useful order of modulation is the second, that is, the frequencies arising from the A_2 term of equation (1). The component of interest here is the term $2A_2PQ \cos [(pt - \theta) - (qt - \phi)]$. As can be seen, this frequency component is proportional to both P and Q and for a given value of P is proportional to Q. In other words the input into the tuned circuit is a linear function of the magnitude of the particular frequency component under consideration in the wave to be analyzed. As will be seen from Table 2, there are a number of other frequency components due to the various orders of modulation in the modulator output. A little consideration, however, will show that a number of these components cannot appear in the output of a resonant circuit tuned to approximately 11,000 c.p.s. where the upper frequency limit of the voice input is 5,000 c.p.s. and the range of the carrier oscillator is limited from the resonant frequency of the tuned circuit to 5,000 c.p.s. above. The only components of Table 2 which can be passed by the tuned circuit are:

$$2A_2PQ \cos [(pt - \theta) - (qt - \varphi)],$$
$$3A_4P^3Q \cos [(pt - \theta) - (qt - \varphi)],$$
$$3A_4PQ^3 \cos [(pt - \theta) - (qt - \varphi)],$$
$$1/2\, A_3Q^3 \cos (3qt - 3\varphi),$$
$$A_4PQ^3 \cos [(pt - \theta) - (3qt - 3\varphi)].$$

In addition to the desired term in the second order modulation of frequency $(pt - qt)/2\pi$ there are two other terms of the same frequency in the fourth order modulation. These have respectively the coefficients $3A_4P^3Q$ and $3A_4PQ^3$. With a given value of carrier input P, the first of these is proportional to Q and it will add to the second order term but will cause no serious trouble. The second term, however, is proportional to Q^3 and, therefore, would cause the input to the tuned circuit to depart from the desired linear relationship with respect to Q. However, the ratio of the cofficient of this fourth order term to that of the second order term is $3A_4Q^2/2A_2$ which is proportional to Q^2 and, therefore, the effect of the fourth order term will fall off rapidly as Q is reduced. In the third order modulation, there is a component of frequency $3qt/2\pi$ which would be passed by the tuned circuit if there were a component of 1/3 the resonant frequency of this circuit in the wave to be analyzed. Considering the extreme sharpness of the tuned circuit, it is rather improbable that such a condition will occur. Moreover, this component is proportional to Q^3 and, therefore, will fall off rapidly as Q is reduced. In the fourth order modulation, there is also a component $A_4PQ^3 \cos [(pt-\theta) - (3qt-3\phi)]$ of frequency $(pt-3qt)/2\pi$. This component would indicate a third harmonic in the unknown wave although it actually contained no other frequency than that of $qt/2\pi$. However, the ratio of its coefficient to the desired second order term is $A_4Q^2/3A_2$ and, therefore, the false indications of a third harmonic can be reduced to any desired extent by reducing Q.

It is evident, therefore, that in an analyzer of this type it is desirable to keep the magnitude of the input of the unknown wave as low as is consistent with obtaining satisfactory meter readings. In the two analyzers now in operation, false indications by the introduction of extraneous frequency components due to the third, fourth and higher orders of modulation are negligibly small. Measurements on an essentially pure frequency within the frequency limits of the apparatus show harmonics less than 0.1 per cent of the fundamental. With an indicated harmonic of not more than this magnitude, it is difficult to tell whether such a harmonic is actually present in the wave or a false indication. At any rate, the harmonic is small enough as to be of no significance. It is, of course, difficult to build the modulator so that it will be exactly balanced over the frequency range covered by the carrier oscillator and, therefore, frequency components such as $(pt-2qt)/2\pi$ appearing in the third order modulation as shown in Table 1 are not totally eliminated. However, the effect of unbalance is of no serious consequence in the practical operation of the analyzer. There is, of course, a possibility of false indications due to higher orders of modulation than the fourth, but the coefficients A_3, A_4, etc., are usually small in comparison with A_2 and in general become successively smaller. Moreover, it will be evident that these false indications may be reduced to any desired extent by reducing the magnitude of the unknown wave.

Analyzer for Complex Electric Waves

By A. G. LANDEEN

IN problems concerned with the electrical transmission of intelligence it is necessary to have means for studying complex electric waves. In certain steady state conditions these complex waves become periodic, and, although not sinusoidal as a whole, may be resolved into a number of sinusoidal components. It is particularly important to be able to measure these components individually.

In studies on systems employing carrier currents which may be transmitted over wire lines it is often necessary to measure a signal wave component which may lie anywhere in the frequency range between 100 and 100,000 cycles per second. The most important range at the present time is, however, below 40,000 cycles per second. In addition to covering a wide range of frequencies these components may also vary considerably in amplitude, both as to absolute value and as to value relative to other components in the signal wave.

For several years there has been in use in the Bell Telephone Laboratories special apparatus by means of which a single component of a complex periodic current wave may be selected from the remaining components and its amplitude determined. The sensitivity and selectivity of this apparatus are such that components of small amplitude may be accurately measured even in the presence of other components of several hundred times the amplitude and differing but little in frequency. With the latest improved form it is now possible to measure current components having amplitudes as low as 10^{-7} amperes with a possible error of 10 per cent. For such minute currents this is within the error which might be introduced by the external apparatus such as attenuators and thermocouples together with their calibration charts.

Though the apparatus was primarily designed for use in current wave analysis work, it may also be readily adapted to voltage analysis. Suitably calibrated, it can be used also as a frequency meter of extremely high precision.

Introduction

The method of analysis here described had its origin in a circuit built by J. W. Horton in 1917. This had a resistance coupled tuned circuit responsive to the component desired. Following the tuned circuit two stages of amplification were used to magnify the selected current. This current was then passed on to a third unit where it was rectified and measured by a D.C. meter. It was evaluated directly by noting the meter deflection and referring to calibrations of the analyzer which had been made with known input currents.

This elementary form of measuring circuit was developed during the World War for the analysis of the sound waves encountered in listening devices used for the detection and location of submarines and torpedoes. It covered the range of audible frequencies and had sufficient sensitivity for its original purpose.

It will be remembered that the first commercial application of multiplex transmission by means of carrier currents came almost simultaneously with the Armistice. The continued study of carrier systems found a useful tool in the current analyzer but placed considerably more rigorous requirements on its performance. These were met by the addition of a second tuned circuit and amplifier system, working from the output of the first, thus giving far greater selectivity than is obtainable in a single circuit. The presence of the multi-stage amplifier between the selective circuits facilitates tuning by avoiding interactions between the circuits. A second modification was the use of a substitution method for evaluating the amplitude of the selected components, as with the considerable increase in the ranges of amplitude and frequency covered, the calibration method for measuring the current became impracticable. To evaluate the current, the output from a sine wave oscillator, which was tuned to the same frequency as the component being measured, was substituted at the input to the analyzer and the amplitude adjusted until it gave the same meter deflection as the unknown component. Since the current from the oscillator is of the same frequency and amplitude as the original component, we can determine the magnitude of the latter by measuring the oscillator output. A convenient means for doing this is to interpose between the oscillator and the analyzer a variable attenuator. It is then possible to fix the oscillator output current at some convenient value, such as 1.0 milliampere or 10 milliamperes and to adjust the input to the analyzer by means of the attenuator. The current can then be read directly from the attenuation tables, it being only necessary to know the location of the decimal point.

The development of the analyzer in this form was carried to the limit of its practicability by F. Mohr. With an analyzer containing three units it was possible to carry through an extensive study of the modulation introduced into the Key West-Havana cable due to the non-linearity of the characteristics of the iron used for loading.

The Heterodyne Method

With the advance in carrier communication, greater refinement in measurement became necessary, calling for still higher selectivity in the analyzer. The best means for accomplishing this appeared to be to heterodyne the wave under investigation in such a manner as to move it to a lower position on the frequency scale. Then with a fixed tuned circuit which would pass only the low frequency current corresponding to the desired component, much greater selectivity might be obtained because of the relatively greater spacing.

To heterodyne the desired component there is required a separate oscillator and a modulator in which the current to be measured and the separately generated current are combined to produce a current of lower frequency. This in effect translates the current under investigation from a high frequency to one of much lower frequency; retaining, however, the relative amplitudes of the components. Since the amount of this translation is determined by the frequency of the local oscillator, a particular component can always be given a certain predetermined value by adjusting the oscillator. This permits the use of a fixed tuned circuit which is highly selective to the difference frequency in the modulator output. By choosing a low value for this frequency it is possible to make the percentage difference between this and interfering frequencies much larger than between the corresponding high frequencies. In the present analyzer 800 cycles per second has been chosen as a suitable value. If, then, the current to be measured had a frequency of 20,000 cycles per second, the local oscillator would be set at 20,800 cycles. It could of course also be set at 19,200 cycles if desired and produce the same difference frequency. If there were also present another current of say 20,500 cycles, the interval in the original wave would only be 2.5 per cent; after heterodyning, however, it would appear as a 300 cycle current, if heterodyned by the 20,800 cycle current. The interval thus becomes nearly 40 per cent of the frequency for which the tuned circuit is adjusted. If these currents are heterodyned directly in a simple modulator, there is also the possibility of modulation between components in the original complex wave. This would result, in the case chosen, in a current having a frequency of 500 cycles, but the percentage difference between the 800 and either the 500 or the 300 cycle currents is many times greater than that between the original high frequency currents so that the fixed tuned circuit would have a very high discrimination to the interfering current.

In some cases the intermodulation between components of the original wave may coincide with the component being measured, or may interfere with the measurement in some other manner. This difficulty is minimized in two ways: first, the amplitude of undesired components is reduced relative to the component under observation, by using selectivity of the type previously described, before impressing the wave on the modulator; second, the modulator is made of a balanced type which permits efficient heterodyne action between the selected component and the heterodyning current, but which gives very little intermodulation between such components as remain after the initial selection. The partial elimination of undesired components before modulation is of further advantage in preventing unnecessary loading of the modulator, which might tend to give it different efficiencies with the complex current and with the sine wave calibrating current.

Description of Apparatus

A schematic diagram showing the several functional elements is given in Fig. 1. The units have been arranged in the order in which

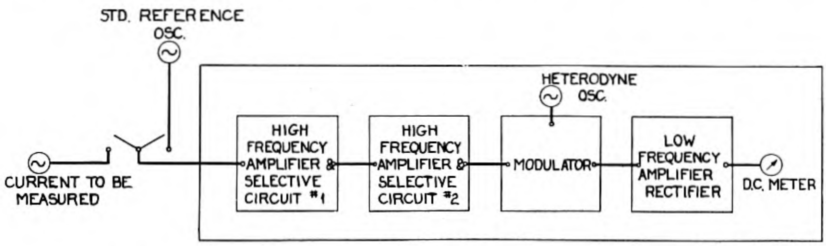


Fig. 1—Schematic diagram of heterodyne current analyzer

the measuring current would proceed, which is also the order of assembly of the completed instrument. In the following, a more detailed description will be given of the individual units.

High Frequency Amplifiers. The circuit for the first unit is shown in Fig. 2 and for the second unit in Fig. 3. These circuits differ mainly in the output terminations of the second tube. The first high frequency unit consists of a simple series tuned circuit together with two amplifier tubes. In the tuned circuit is also included the coupling resistance, R, for controlling the input to the amplifier. The sharpness of resonance will therefore depend to some extent upon the value of this resistance, but under most operating conditions it is relatively small in comparison to the total effective resistance which includes that of one tuning coil and the condensers.

In both of these circuits the A.C. input voltage to the first tubes is obtained from the drop across the condensers. If it is desired to discriminate against a component of higher frequency than the one being measured, it is advantageous to use the voltage across the condenser, whereas if the interfering component is of lower frequency, the voltage across the inductance should be used. It is therefore desirable to make

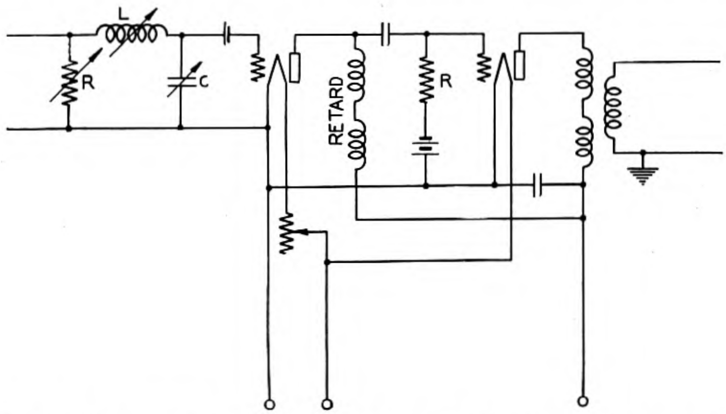


Fig. 2—High frequency amplifier unit No. 1 of heterodyne current analyzer

provision by means of a switching arrangement whereby the voltage may be applied from either the inductance or the capacity depending upon the discrimination desired. With the coils and condensers used the voltage across the condenser at resonance is over a hundred times that across the coupling resistance.

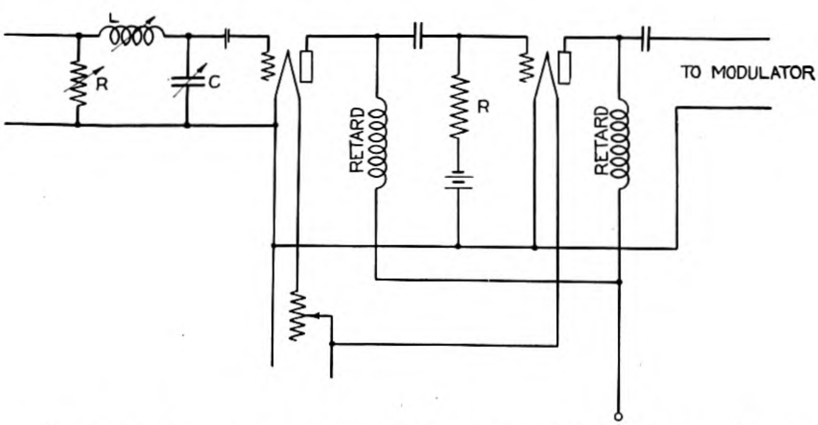


Fig. 3—High frequency amplifier unit No. 2 of heterodyne current analyzer

The first tube of each unit serves as a voltage amplifier. This works into a high resistance between the grid and filament of the power tube.

Because of the wide range of frequency over which the amplifiers will be operated, resistance, rather than transformer coupling, was chosen as the most reliable form of interstage connection. Due to the large step-down in impedance necessary between the first amplifier and the coupling resistance of the second selective circuit, a transformer having

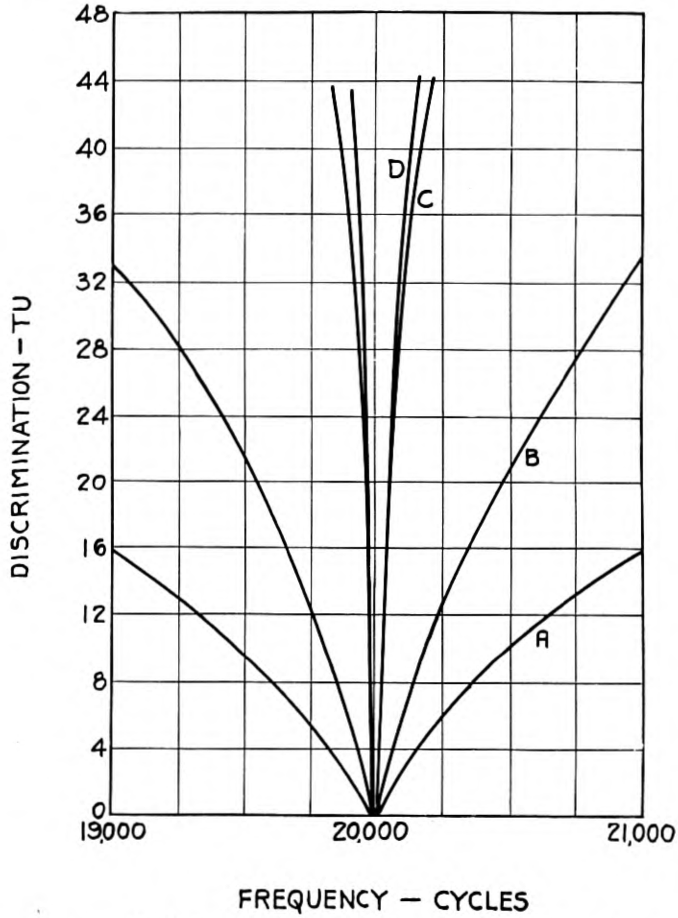


Fig. 4—Discrimination curves of heterodyne current analyzer

a high step-down ratio was used. The output of the second amplifier also works into a transformer through a fixed resistance and potentiometer as shown in Fig. 5. The purpose of the fixed resistance is to maintain a more uniform output impedance as the potentiometer is varied. Though the two amplifiers are almost identical in their circuits, they differ in the plate voltage. The first section is operated at

240 volts since its tubes are subjected to the heaviest input, being preceded by only one selective circuit which can but partially eliminate large interfering currents. This seems a most unusual arrangement until it is remembered that, although the amplitude of a particular part of the current may be increased, the total load on the first stage may well be greater than that on the final stage. This is the reverse of the situation in cascade amplification where the first tubes handle only a small current and the succeeding ones a proportionately larger current.

The discrimination of one of the tuned circuits when a coupling resistance of 1 ohm is used is given by curve *A* of Fig. 4. Curve *B* shows the effect of adding the second tuned circuit. If regenerative amplification were employed, both the selectivity and amplification could of course be greatly increased, but this has not been used because of the necessity for high stability and measurement precision.

Modulator. As previously mentioned, the second amplifier unit works into a modulator, the circuit of which is shown in Fig. 5. This is

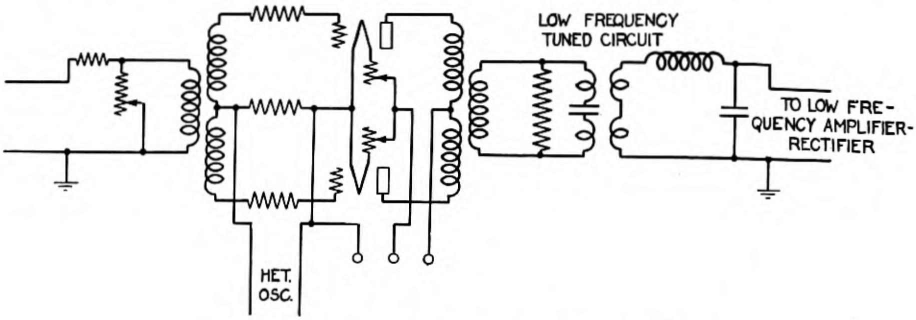


Fig. 5—Modulator for heterodyne current analyzer

of a two tube balanced type in which modulation, or frequency transformation, takes place in the grid circuit. The heterodyning frequency is applied in the common input lead across a suitable resistance. The input from the amplifier is applied through a transformer across the grids of the two tubes in series with a high resistance in each side. No biasing potential is applied on the grids. A modulator operated in this manner has the property of giving a modulation output proportional to the smaller of the two input currents and independent of the larger. The amplitude of this output may, therefore, be determined entirely by the amplitude of the component being measured. Another desirable characteristic of this type of modulator is that its efficiency is not affected by interference, hence it will show a fixed relation between the low frequency output and a given input component regardless of the presence of other interfering currents in the input side. Through the use of a balanced circuit intermodulation is reduced considerably below the limit possible with a single tube modulator.

The output of the modulator is connected directly to a double tuned circuit which selects the low frequency modulation product corresponding to the component of the complex wave being examined. The frequency to which this circuit is adjusted is, as already mentioned, 800 cycles.

Low Frequency Amplifier-Rectifier. The 800 cycle output from the modulator is generally too small to measure on a meter of the usual type without first being amplified. For this reason a low frequency amplifier has been added and the output of this rectified so that all measurements could be made on a sensitive D.C. meter, having a full scale deflection of 1 or 2 milliamperes.

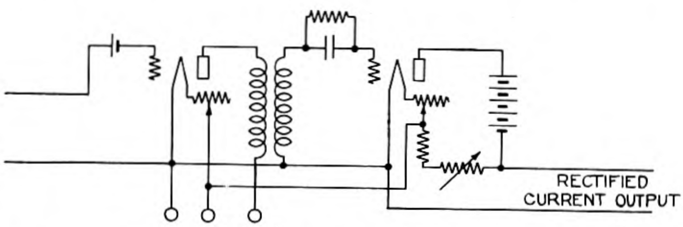


Fig. 6—Low frequency amplifier-rectifier for heterodyne current analyzer

The combined low frequency amplifier-rectifier circuit is shown in Fig. 6. A step-up transformer is used between the amplifier tube and the rectifier in order to increase the amplification so that, for a given A.C. output, a smaller input to the first amplifier might be used. Since this circuit is always to be operated at the one frequency (800 cycles), its overall frequency range characteristic is not of particular interest and its performance was studied only at the one frequency. The rectifier is of the grid leak and condenser type. Its performance differs from the usual type of rectifier since it is operated over that portion of its characteristic which gives a linear relation between input voltage and direct current output. By suppressing the space current corresponding to zero input the accuracy with which data can be taken is greatly increased. As shown by the circuit arrangement in the output side, part of the "A" battery current is used to oppose the space current through the meter. This permits using a meter of high sensitivity, having a full scale deflection of one or two milliamperes, on which 1/100th part of a milliampere can easily be read.

Heterodyne Oscillator. The major requirement to consider in the design of the heterodyne oscillator was that of frequency stability. As the only function of the oscillator was to furnish a current for heterodyning the one being measured the output requirements were moderate, 10 milliamperes into 600 ohms being ample, but it was important that the frequency remain constant during a series of measurements as even slight variations, of a fraction of a per cent, would change the attenuation of the 800 cycle tuned circuit to the sideband current. Stability in this case depended mainly upon the "A" and "B" battery voltages, since the output load consisted of a pure resistance and there was no reaction back to the oscillator due to a variable output impedance.

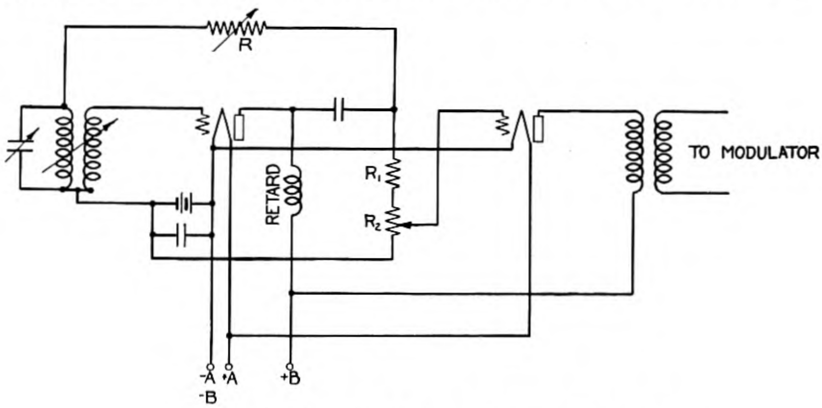


Fig. 7—Oscillator for heterodyne current analyzer

The oscillator circuit is shown in Fig. 7. This shows two tubes, one as an oscillator and one as an amplifier. The coupling consists of a 20,000-ohm resistance used as a potentiometer, which is placed in series with a 100,000-ohm resistance and the two used as the oscillator load. This makes the coupling impedance only one sixth of the total oscillator output impedance and therefore reduces the effect which the amplifier tube might have on the frequency. The change in frequency due to the "A" and "B" voltage can also be controlled by inserting a high resistance in the feed-back path between the plate and oscillation circuits. This should be several times that of the tube impedance so that any change in the latter would then be a proportionately smaller part of the total impedance and hence have a less effect upon the frequency.

The selection of tuning coils for various frequency ranges is made by keys which at the same time select the proper feed-back resistance. Only three coils are used to cover the frequency range between 3,000 and 50,000 cycles. The output of the amplifier tube works into the resistance in the mid-branch of the modulator input circuit.

Mechanical Construction

Figs. 8 and 9 show the front view of the complete analyzer as in present use and also the interior view of an individual unit (the heterodyne oscillator), to indicate the method of construction.

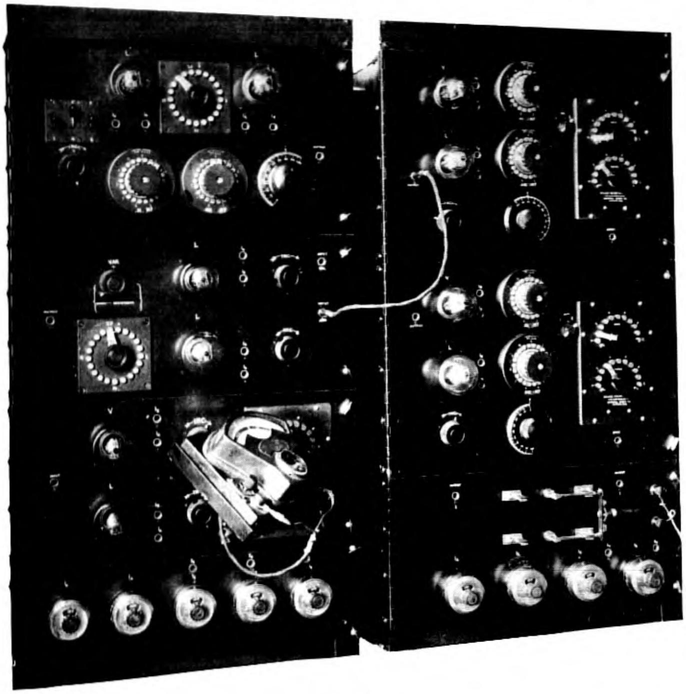


Fig. 8

Each section is completely enclosed in metal, by having shielded cases in the rear and heavy metal panels in front. Perhaps the feature of most interest in the mechanical construction is that of the hinged front panels. These were made of 1/4 in. aluminum to insure ample strength and rigidity with a minimum of strain on the supporting hinge. Each panel is provided with two thumb screws on the edge opposite the hinge so that the units may be readily inspected. This feature of accessibility is particularly desirable in making periodic inspections and in renewing the "C" batteries, which have been mounted on the panels. The heavy material such as tuning coils has been mounted inside the case. The flexible leads, as shown in the lower corner of the opened unit, connect the panel apparatus with the "A" and "B" batteries, and with the modulator. All of the panels are provided with a metal strip, such as shown along the top and bottom edge, which fits into a groove in the case and thereby provides better shielding between

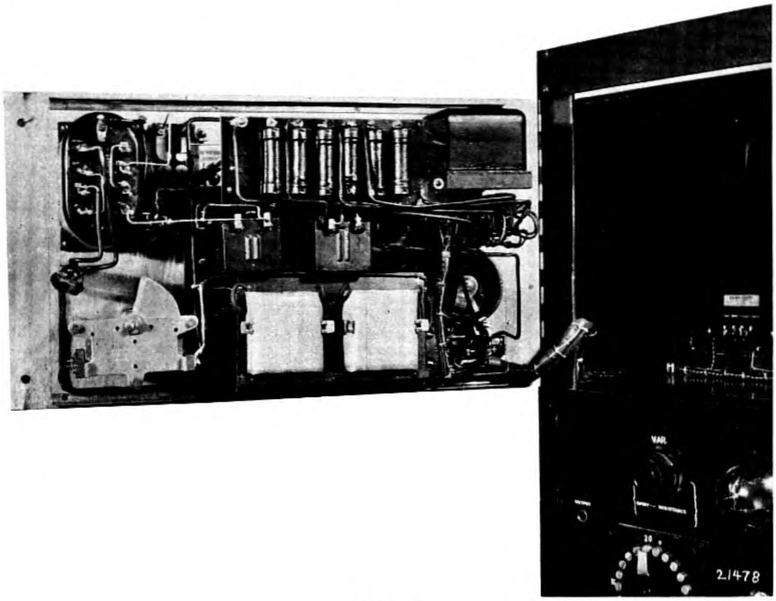


Fig. 9

each of the units. Since each unit of the analyzer is provided with a separate shielding case, two sheet iron walls are interposed between any pair, forming a double shielding to electric fields occurring within the analyzer. The complete unit of two sections is very economical in space as the overall height, width and depth of each section are only 36, 19 and 12 1/2 inches respectively.

Characteristics of Complete Current Analyzer

The successful operation of the heterodyne current analyzer depends, of course, upon knowing its limitations and its reliability. Under limitations may be grouped sensitivity, selectivity and modulation in the analyzer itself; and under reliability, the limits within which readings can be repeated.

Sensitivity. The sensitivity depends upon the coupling resistance used in the two amplifier units, and increases with added resistance so long as this is only a few ohms and small as compared to the total effective resistance in the selective circuits. When these are made 1 ohm each, it requires only 1 microampere to give one milliampere of rectified current. With 10 ohms coupling resistance only 10^{-8} amperes input would be required. This, however, is a larger coupling than it is desired to use, since the analyzer becomes too sensitive and susceptible to mechanical vibrations as well as to electrical interference from outside sources.

One desirable feature is that the sensitivity characteristic of the complete current analyzer is a straight line so that doubling the input will give twice the deflection in the meter reading the rectified current. This is an advantage since, if the deflections and input amplitudes do not change by the same ratio, the presence of interfering currents is indicated.

Selectivity. The selectivity of the current analyzer depends upon the time constants of both the high and low frequency tuned circuits, and to some extent upon the coupling resistances, but the latter are usually relatively small and do not have an appreciable effect. The discrimination obtained by the use of the heterodyne method and fixed low frequency selective circuit is shown by curve C of Fig. 4. This curve may be compared with curve A, which shows what can be done with high grade elements in a single tuned circuit. The discrimination of the complete analyzer, including the initial stages and the heterodyne stage, is given by curve D. Tests with two frequencies show that if one is 250 times as large as the other the smaller may be measured without appreciable error if the difference in frequency is not less than 1 per cent; if the ratio of amplitudes is 1,000 to 1, the frequency difference need not be less than 2 per cent.

Modulation. As to modulation in the current analyzer there are two sources which contribute, the vacuum tubes and the tuning coils and transformers. Of these the tubes are the most troublesome, since they furnish both even and odd order modulation products, whereas the coils contribute only to the odd orders. Of these the third is generally the only one that is of any interest since higher odd orders are too small to produce any interference. Modulation need be considered only when measurements are made of small components in the presence of very large ones, as it is under these circumstances that conditions for modulation are the most favorable. This condition requires high sensitivity which is obtained by increasing the coupling resistance, and large resistance means greater interference voltage on the first amplifier and also less selectivity in the tuned circuits. The result is an increased load on all sections of the current analyzer, which causes modulation and may produce an error in the readings. It is, therefore, important to know the limitations which this imposes upon it as a measuring device. Then with this information data can be taken within known limits of accuracy. Measurements have been

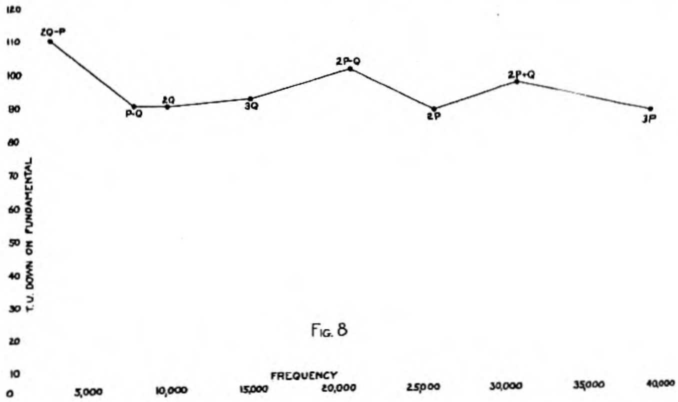


Fig. 10—Second and third order modulation in heterodyne current analyzer together with 5,000 and 13,000 cycle band-pass filters with one or two input frequencies

made on the combined modulation occurring in the analyzer and also in the filters which were necessarily associated with the measurements. These were made with one input current and also with two input currents of different frequencies, which were applied simultaneously to the analyzer. One or two filters were therefore necessary to suppress all components except the fundamental currents desired, but any small amount of modulation which would occur in the filters would add to that produced in the analyzer so that the results shown by the curve in Fig. 10 represent the total modulation in both the filters and the analyzer. The modulation amplitude is expressed in terms of transmission units with respect to the current into the analyzer. This curve is quite irregular and depends somewhat upon the frequency. It is the lowest at 39,000 cycles where the modulation current is shown to be 86 *TU*[1] down or 0.00005 as large as the current into the analyzer. Measurements can, therefore, be made at this frequency of the modulation occurring in any device when its amplitude is not less than 66 *TU* below the amplitude of the fundamental, with a possible maximum error of 10 per cent. At other frequencies this amplitude may be less, as for instance at 21,000 cycles, measurements may be made up to 80

[1] For a discussion of this method of expressing current ratios see "The Transmission Unit and Telephone Reference Systems" by W. H. Martin, *Journal A. I. E. E.*, June 1924, Vol. XLIII, No. 6; also *Bell System Technical Journal*, July 1924, Vol. III, No. 3; also "The Transmission Unit," R. V. L. Hartley, *Electrical Communication*, July 1924, Vol. III, No. 1.

TU without exceeding the same percentage of error, or up to 60 *TU* with 1 per cent error due to undesirable modulation in the current analyzer.

Reliability. Use of the heterodyne current analyzer over a period of two years has proven it to be one of the most reliable means for making measurements. With proper maintenance, which consists only in maintaining constant "A" and "B" battery voltages and grid voltage, and with proper precautions as to shielding and balance, readings can be taken with a precision of 2 per cent.

Vacuum Tube Curves Obtained with Heterodyne Current Analyzer. A number of curves have been added to illustrate the application of the current analyzer, though of course these represent only a small part of the field of usefulness for which it is adapted.

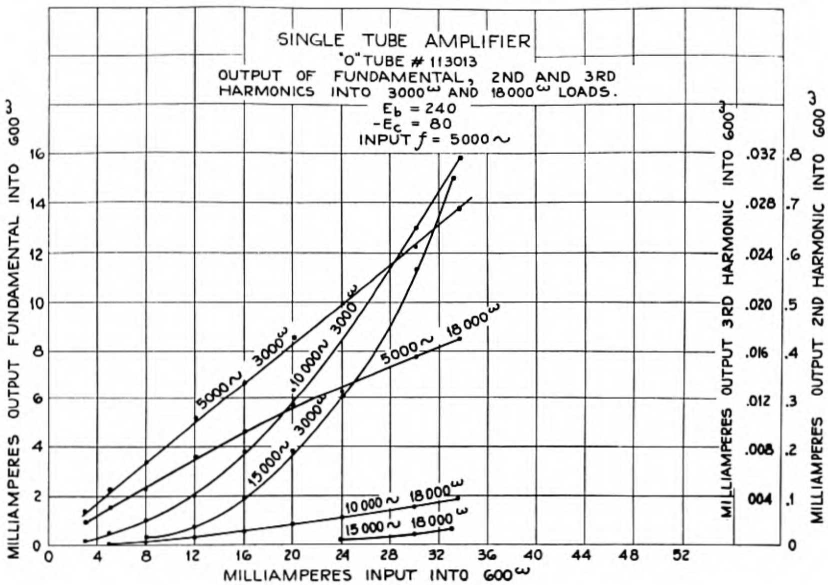


Fig. 11

The first set of curves, shown in Fig. 11, were taken on an "O" tube (104–D) to show how the fundamental current and the second and third harmonics produced in the tube changed with increase in the input amplitude of a single frequency. Two sets of curves are shown which were taken for two values of load impedance, one being equivalent to the normal tube impedance and the other being six times as large. The output currents have all been computed to show the equivalent output into 600 ohms which is a common reference standard of impedance used in telephone work.

The second set of curves, shown in Fig. 12, were taken with an "L" tube (101–D) but with two input frequencies of $Q = 5{,}000$ and $P = 13{,}000$ cycles applied simultaneously. The measurements in

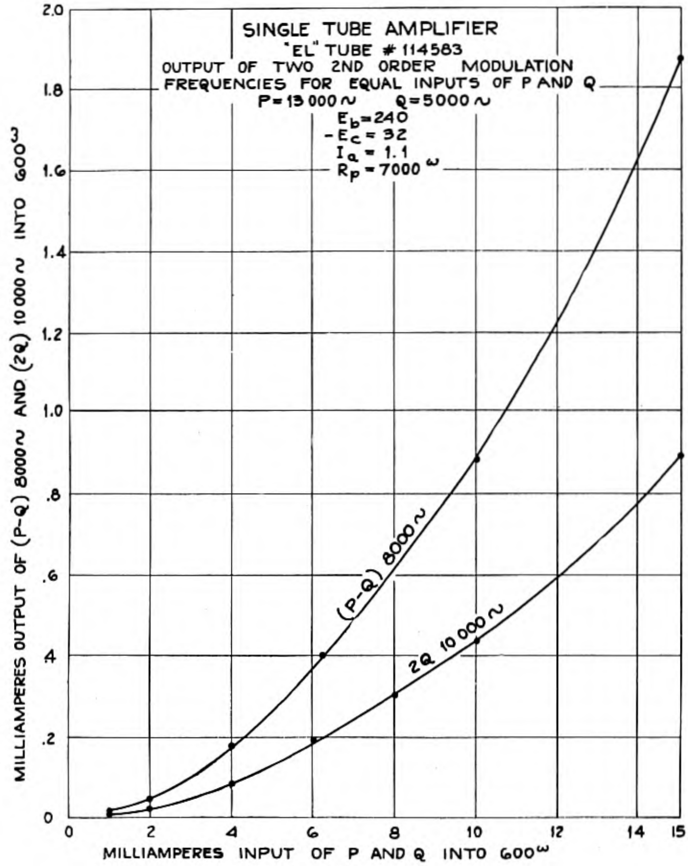


Fig. 12

this case were made on the modulation product consisting of the difference frequency $(P - Q) = 8{,}000$ cycles and of the second harmonic of Q, $2Q = 10{,}000$ cycles. Such curves furnish a quick and accurate means of studying the performance of the vacuum tube and also afford a convenient check on mathematical computations which, in this particular case, with a two frequency input, have indicated that the difference frequency amplitude should be twice that of the second harmonic.

In Fig. 13 are shown the third harmonic output of an "O" tube with a single input frequency of constant amplitude and with a variable load impedance. The harmonic current shown by this curve could

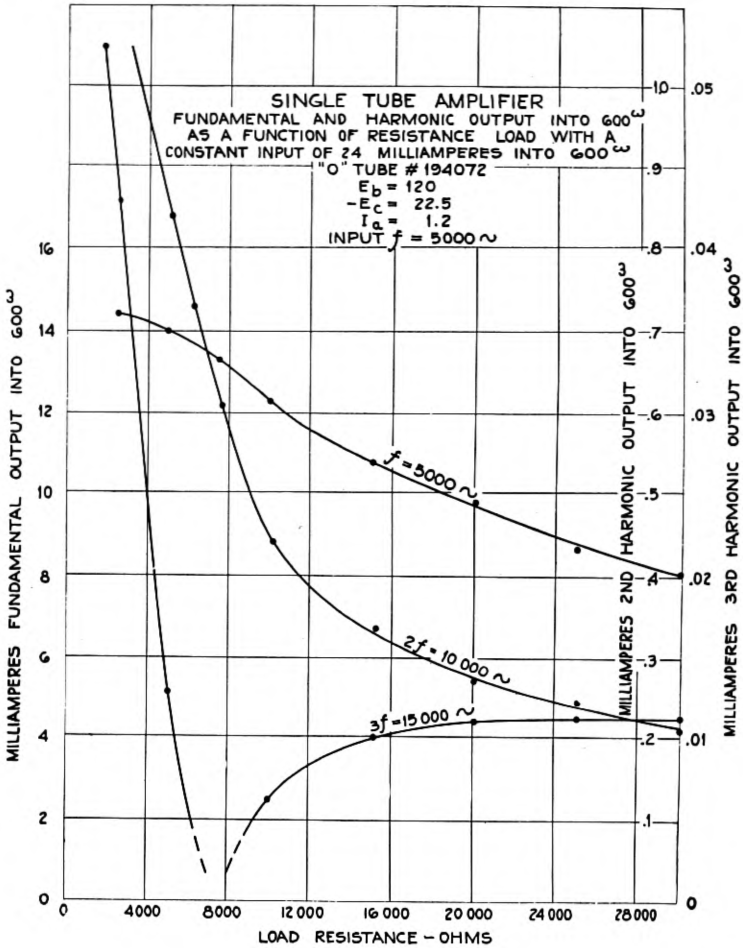


Fig. 13

not easily be predicted from computation due to the many factors involved, and can be measured only by an instrument of high sensitivity that can faithfully follow the varying amplitudes of a minute current.

Voltage Analyzer. In addition to its use as a current measuring device the current analyzer may be used, as previously mentioned, for measuring the voltage in circuits where the power dissipation is very small, or where the voltage cannot be detected except by several stages of amplification such as are obtained in the current analyzer. To adapt it for this purpose it is only necessary to precede it by a simple circuit such as shown in Fig. 14. This consists of a single

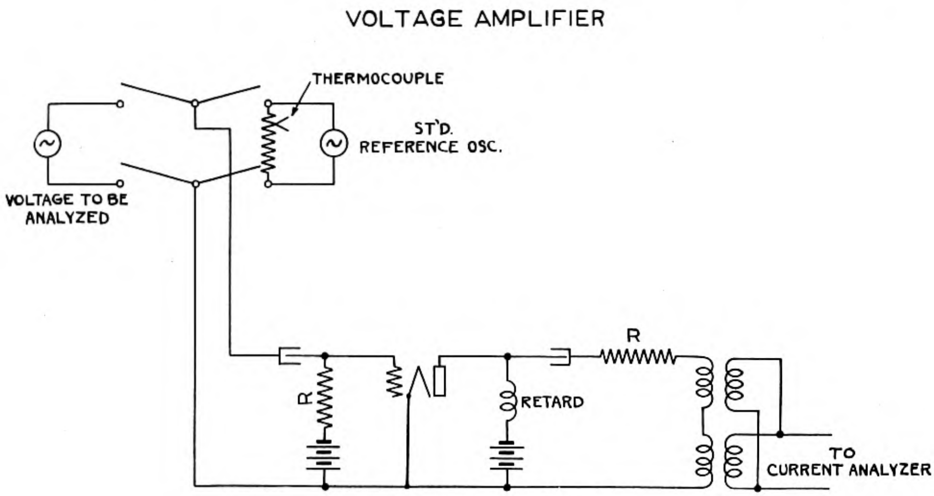


Fig. 14—Voltage amplifier

vacuum tube having a large resistance across the grid and filament. This resistance should be greater than the impedance across which the voltage is to be measured. The output side works into a step-down transformer through a resistance of several times the output impedance of the tube. This tends to straighten out the characteristic and to lower the tube modulation level. The object of the step-down transformer is of course to secure greater efficiency in working into the low input impedance of the heterodyne current analyzer.

The measuring procedure would be to apply the voltage to be evaluated across the high impedance input and adjust the analyzer in the usual manner; then substitute the output from the standard oscillator of the same frequency across the amplifier and adjust the amplitude to give the same meter deflection. In order to determine the voltage applied, the oscillator may be connected across a known resistance in parallel with the input, and the current into this resistance measured. The IR drop will then be a measure of the voltage applied.

The range of voltage which can be measured of course depends upon the biasing potential on the amplifier grid as it is not desirable that grid current flow through the high resistance and increase the tube modulation. The most frequent use of the amplifier in the laboratory has been in measuring voltages around 10^{-2} volts but it can also be equally well used to measure much smaller values, of the order of 10^{-4} volts, when the frequency employed does not make the input impedance of the first tube too low.

In the choice and arrangement of the elements of this system and for many of the details of its adjustment recognition is particularly due to the extensive contributions of E. Peterson, W. A. Mueller and C. R. Keith.

Transatlantic Radio Telephony

By RALPH BOWN

MANY of the technical and scientific features of Transatlantic Radio Telephony have been discussed individually in considerable detail in engineering papers. Furthermore, through the agency of the newspapers much general information has been published regarding the development of commercial telephone service between the old world and the new. Most of this published material either is sketchy in nature or is concentrated upon some detail of the system and it is difficult to gain from it a connected picture of how the final result was built up through several years of continued effort. The following has been written in an attempt to provide such a connected story.

As soon as the successful experiments carried out by the Bell System engineers in 1915 had resulted in the reception of intelligible speech in Paris and Honolulu transmitted from near Washington, D. C., it became a foregone conclusion that sooner or later a serious attempt would be made to bridge the Atlantic Ocean by radio telephone service which would be available to the public at large.

While the 1915 experiments were successful, they also served to emphasize the tremendous difficulties which had to be overcome. The onset of war activities prevented continuing a direct attack on these difficulties but the developments incidental to the wartime use of radio had a profound effect on the instrumentalities necessary to their solution. In particular the development of vacuum tubes for transmitting purposes made considerable progress. Other radio developments carried out immediately subsequent to the war also aided the program.

When transatlantic telephony was taken up again for active consideration, it was obvious that the first requirement was for a transmitting station which would be sufficiently powerful to deliver satisfactory signals on the other side of the ocean. Since the amount of power which would be required to do this was unknown, it was decided to construct a transmitter which was sufficiently large to approach the economic limit of what it seemed it could be worth while to attempt to employ in a commercial undertaking. For this purpose there were available water-cooled vacuum tubes[1] each capable of handling about

[1] "A New Type of High-Power Vacuum Tube," W. Wilson: BELL SYSTEM TECHNICAL JOURNAL, Vol. 1, No. 1, July 1922, pp. 4–17.

10 kw. of power. It was decided that about 20 of these tubes were as many as could be reasonably expected to work satisfactorily in a parallel combination. In order to use these powerful tubes in the most advantageous and economical way, the transmitter was constructed to radiate what is called a single sideband carrier eliminated transmission.

In the ordinary radio telephone transmission such as is used in broadcasting, the radiation sent out consists of a carrier frequency together with two sidebands. The carrier transmits no intelligence but the complete message is transmitted in duplicate since each sideband contains the entire message. By eliminating one of the sidebands and the carrier it is possible to send out the intelligence using only one sideband. If the entire power capacity of the transmitting system is thus concentrated on a single sideband, the power is used several times more effectively.

Since it is more difficult to filter a single sideband away from its carrier and its brother sideband as the frequency becomes higher it was decided to produce the single sideband at a relatively lower frequency as is done in wire carrier telephony and then step it up by a modulation process to the desired position in the frequency range.[2] The voice was therefore modulated upon a 30-kilocycle carrier, and the single sideband produced by passing the modulated result through a band pass filter. This band is then combined in a second modulator with a frequency of 90 kilocycles and the resulting difference frequency, which is a sideband at 60 kilocycles, after passing through another band filter is ready to be amplified to high power for radiation from the antenna. Four preliminary stages of amplification are necessary before the final high power 20-tube amplifier is reached.[3]

When this transmitting apparatus first became available for experimental trial, it was set up at the large radio station at Rocky Point, Long Island, since the experiments were at that time being made in cooperation with the Radio Corporation of America, and the Radio Corporation arranged to lend one of its large and efficient antennas [4] at that station. Subsequently this antenna was leased for use in the final experiments and in giving a commercial service.

[2] "Production of Single Sideband for Transatlantic Radio Telephony," R. A. Heising: *Proceedings of the Institute of Radio Engineers*, Vol. 13, No. 3, June 1925, pp. 291–312.

[3] "Power Amplifiers in Transatlantic Radio Telephony," A. A. Oswald and J. C. Schelleng: *Proceedings of the Institute of Radio Engineers*, Vol. 13, No. 3, June 1925, pp. 313–361.

[4] For a description of this type of antenna known as a multiple tuned antenna see "Transatlantic Radio Communication," E. F. W. Alexanderson: *Proceedings of the American Institute of Electrical Engineers*, Vol. XXXVIII, Part II, 1919, p. 1089.

Simultaneously with the development of this transmitting apparatus, the art of measuring the strength of received radio signals and the amount of static, or radio noise, present at a receiving station had been developed.[5] Therefore in order to try out the effectiveness of the transmitting apparatus, engineers provided with suitable measuring equipment were dispatched to England and set up their apparatus near London. Satisfactory signals were received from the Rocky Point transmitter and in January 1923, it was possible to demonstrate one-way talking across the Atlantic Ocean on a much more satisfactory basis than had previously been possible.

Then there ensued a program of weekly tests wherein signals were sent from Rocky Point each hour for the 24 hours of one day each week and measurements of received signals, radio noise, and intelligibility tests of spoken words were made in England. This one-way telephone circuit was in other words used as a sample whereby the variations to which radio telephony is subject could be explored, catalogued and studied over an extended period of time so estimates could be made of the improvements which would be necessary before anything in the way of reliable communication could be established.

The British General Post Office became so interested in the subject as the result of the initial experiments that they decided to cooperate with the American Company to the fullest extent in determining what the possibilities of transatlantic radio telephony were. They therefore constructed an experimental receiving station and made arrangements to have a transmitter similar in general character to that being used at Rocky Point installed in the new high-powered telegraph station then under construction at Rugby.[6]

The study of transmission initiated in 1923 has been continued to the present time and a large volume of statistical information has been collected.[7] There are two main kinds of variation which have to be contended with. First, the strength of signal changes radically

[5] "Radio Transmission Measurements," Ralph Bown, C. R. Englund and H. T. Friis: *Proceedings of the Institute of Radio Engineers*, Vol. 11, No. 2, April 1923, pp. 115–152.

[6] "The Rugby Radio Station of the British Post Office," E. H. Shaughnessy: *Journal of the Institution of Electrical Engineers*, Vol. 64, June 1926, pp. 683–713. Also "Transatlantic Radio Telephony. Radio Station of the British Post Office at Rugby," E. M. Deloraine: *Electrical Communication*, Vol. 5, July 1926, pp. 3–21.

[7] "Transatlantic Radio Telephony," H. D. Arnold and Lloyd Espenschied: Bell System Technical Journal, Vol. II, No. 4, pp. 116–144, or *Journal of the American Institute of Electrical Engineers*, August 1923. Also "Transatlantic Radio Telephone Transmission," Lloyd Espenschied, C. N. Anderson and Austin Bailey: Bell System Technical Journal, Vol. IV, No. 3, pp. 459–507, or *Proceedings of the Institute of Radio Engineers*, Vol. 14, No. 1, February 1926, pp. 7–56.

with the time of day, being stronger at night. Second, the amount of radio noise present is usually less in the morning and increases towards the evening and well into the night. It is not the absolute strength of the signal which is controlling, but the extent to which it dominates the noise, therefore the ratio between the signal and the noise is the thing which indicates the satisfactoriness with which communication can be carried on. While signal transmission does not change widely between summer and winter, the amount of noise present in the summer time is usually very much greater than that in the winter time, so that the difficulties of communication in the summer are greatly increased.

It soon became apparent that the amount of increase in the signal-to-noise ratio which would be necessary to put conversation on anything like a practical basis would be so great that to hope to get it by increasing the power at the transmitting end was quite out of the question. Thus improvement had to be looked for at the receiving end and the problem became one not of increasing the signal strength, but of decreasing the amount of noise which was allowed to get into the receiving set along with the signal. There are three known ways of decreasing the effect of static in a case of this kind.

Since static is distributed over the entire frequency range, the first and most obvious one is to use such high selectivity that only the signal frequencies are permitted to enter the receiving set. This reduces the amount of static to that which is encompassed by the frequency band occupied by the signal.[8] If suitable band filters are employed it is possible to obtain a degree of selectivity such that practically all static noise which can be eliminated by this method is prevented.

A second method of reducing the amount of static is to employ receiving antenna systems which are directional, in other words, systems which are receptive only to signals coming from the direction of the transmitting station and are blind to interfering signals or interfering static coming from other directions. The most practical system which has so far been developed for doing this at long wave-lengths is the so-called wave antenna.[9] This consists of an open wire line three or four miles long which is grounded at both ends in the characteristic impedance of the wire-to-ground circuit. Thus it is substantially an aperiodic system, there being no reflections at the terminals. Radio waves which approach this line from the side produce relatively very

[8] "Selective Circuits and Static Interference," John R. Carson: BELL SYSTEM TECHNICAL JOURNAL, Vol. IV. No. 2, April 1925, pp. 265–279.

[9] "Wave Antennæ," H. H. Beverage, Chester W. Rice and E. W, Kellogg: *Journal of the American Institute of Electrical Engineers*, Vol. 42, March 1923, pp. 258–269; May 1923, pp. 510–519; and July 1923, pp. 728–738.

small currents in the grounding wires at the terminals. Waves which approach longitudinally build up currents in the line wire as they proceed along it and cause relatively large currents in the grounding wire at the distant end. By building such a line so that it points towards the transmitting station and attaching the receiving set at the end most distant from the transmitter it is possible to obtain a considerable degree of directivity. Further development of this simple basic antenna system is obtained by attaching to it various balancing arrangements. Or by combining several antennas together, it is possible still further to improve the directional characteristics.

In order to determine how much advantage could be obtained by the use of wave antenna systems, the British Post Office constructed one at their receiving station in England and the use of one at Riverhead, L. I. was borrowed by the telephone company from the Radio Corporation of America. Measurements taken over a considerable period of time showed that the signal-to-noise ratio on a good system of combined wave antennas was about ten times as great as that on a simple loop antenna. This was very gratifying since it meant that by the use of wave antennas the transmission would be improved just as much as if the transmitting station power had been multiplied 100 times.

At the receiving end quite aside from the special nature of the directive antenna systems, the amplifying and detecting apparatus must be of a character suited to the amplification and detection of single sideband-carrier eliminated signals. In order to detect signals of this kind it is necessary that a carrier be resupplied to the signal before detection. This is done by means of a local oscillator. Thus the signal actually supplied to the detector differs from the ordinary signals such, for instance, as are used in broadcasting, only by the fact that one of the sidebands is missing. This is of no importance since the complete signal may be detected from the carrier and one of the sidebands. The actual apparatus being used at the American receiving station is similar to the modulating apparatus employed at the transmitting end for producing a single sideband in that the process at the receiving end is substantially reversed. By means of a double detection type receiving set having a beating oscillator frequency of about 90 kc., the 60 kc. incoming sideband is reduced to approximately 30 kc. in the first detection. It is then passed through filters, amplified and has added to it the carrier frequency. The second detection brings it back to voice frequency and after further amplification it is ready to go on to the wire line to the terminal.

A third way in which it is possible to avoid the effects of static is to place the receiving station in a more northerly latitude, bringing the signals down to the business centers by wire. In order to determine the extent to which this was useful, measurements were made at Green Harbor, Mass., and at Belfast, Me., over a considerable period of time. These measurements were made on telegraph signals specially transmitted by the British General Post Office from its telegraph stations. It was found that in Maine the signal-to-noise ratio was, at least during the important hours of the day, something like six or eight times that which was obtained on Long Island. In other words, the improvement was as great as would have been obtained by multiplying the power of the English transmitting station some fifty times. It was therefore decided to build a receiving station equipped with wave antennas at Houlton, Me.

These two improvements, the one in the antenna and the other in its location, taken together comprise an astounding advance in the battle against static. To get the same results by increasing the power of the transmitting station while using older receiving methods it would be necessary to employ transmitting apparatus rated at one million kilowatts, a power which is obviously far beyond either the technical or economic possibilities.

The British Post Office, having in mind that all Great Britain was already more northerly in latitude than Maine, decided to build a temporary receiving station near Wroughton, England, leaving the question of a more northern location for later experiments.

On both sides of the Atlantic, suitable wire circuits had been arranged to tie the transmitting and receiving stations, to terminal points in New York and London. There were then available early in 1926 the means whereby a complete channel could be set up from New York to London and one from London to New York. These two channels were operated on different radio frequencies, the American transmitter sending on 57 kc. while the British transmitter sent at about 52 kc.

At this point, with the major radio problems if not solved, at least well in hand, the undertaking became more a telephone toll circuit problem for the time being. The simplest way to connect up a system of this kind is to follow the practice which is employed for long 4-wire telephone circuits. Where the circuit needs to become a 2-wire circuit for termination in an office where it may be switched to subscribers, the outgoing and incoming wires are brought together through a hybrid coil or 3-winding transformer. This well-known device, by means of a balancing arrangement, has the property of directing currents incoming on the receiving leg of the 4-wire circuit into the

2-wire line without permitting them to go into the outgoing leg of the 4-wire circuit. The currents coming from the 2-wire line go into both sides of the 4-wire circuit but travel on the receiving leg only until they meet with a repeater which, being directed against them, prevents further travel. The amount of amplification which can be maintained in such a circuit is dependent upon the effectiveness of the balance maintained between the real 2-wire line and the artificial line or network at the hybrid coil. The transatlantic circuit was set up for initial two-way experiments in accordance with this procedure. Since the east-bound and west-bound channels were on different frequencies, the selectivity of the receiving sets prevented any cross-fire from the local transmitter into the local receiving circuit.

Since it is necessary to deliver signals to the distant receiving station of the maximum possible amplitude in order to maintain a favorable signal-to-noise ratio, it was essential that the transmitters be kept loaded up to full output even though the voice currents coming from the speakers might vary widely due to differences in voices and differences in attenuation of connected 2-wire circuits. This was done by changing the gain in the repeaters, the operation being carried out by a control operator in a manner similar to that employed in broadcasting stations. In order to maintain the overall gain around the circuit constant to avoid singing difficulties, it was necessary to change the amplification at the receiving end in such a manner as to compensate for changes at the transmitting end.

Experimental operation of the system on this basis was hindered by the fact that the two frequency bands being employed for the two oppositely directed channels were also being used by a number of radio telegraph stations, some of these being so powerful as to produce interfering signals which very seriously hampered telephone conversation. It was evident that some arrangement must be made to enable the telephone communications to be carried on in frequency bands which were used by them exclusively. The fact that radio telephony inherently requires a wider band for its accomplishment than does radio telegraphy made it desirable to use every device available to narrow the band occupied in order to reduce to a minimum the necessary displacement of existing telegraph services. The employment of single sideband carrier eliminated transmission had already cut in half the frequency space required over that which would be needed if the ordinary form of modulated transmission were used. In order to cut down still further the width of frequency band occupied it was decided to attempt to operate both the east-bound and the west-bound channels on exactly the same frequency band. If this could be done the entire system would utilize only about 3000 cycles.

In this sort of arrangement, it is evident at once that selectivity at the receiving station is of no further avail in preventing interference from the local transmitter and that unless means are provided greatly to reduce this crossfire or to set up the circuit in some fashion so that it is harmless, the local circuit from transmitter to receiver with return by wire will be in a singing condition, since it is not practicable to obtain at the hybrid coil anything like a sufficient balance to prevent this.

It was found that with the American receiver in Maine some 500 miles from the transmitting station, the local signals were so reduced by distance that the further reduction which could be obtained by virtue of the antenna directional characteristics was sufficient to permit operation. At the English end, however, due to the proximity of the receiving station to the transmitting station, this so-called "radio balance" method of operating could not be employed. This difficulty had been foreseen and there had been developed a switching device based upon certain similar switching devices called echo suppressors which are employed in long toll circuits.[10] The function of this apparatus was in part to supplement the hybrid coil in its office of preventing received signals from getting into the transmitting line. The arrangement is one in which switching means are employed alternately to disable the transmitting or receiving side of the radio circuit automatically in response to the voice currents produced by the speakers at the two ends. Each end of the system was provided with a device of this character operating on substantially the same principles. Briefly, the functioning of the device is as follows:

When no one is speaking on the circuit the transmitting voice paths are blocked at both the New York and London ends of the system but the receiving paths are open so that incoming radio signals pass freely through to the ears of the subscribers. When a speaker, for instance, in America, speaks, his voice currents actuate the device to block off his receiving path and to open his transmitting path so that his voice goes out. Since the other end of the circuit is in a receiving condition, the voice currents travel through the entire system to the listener's ear. When the American speaker has finished, his apparatus is automatically restored to the receiving condition and the British speaker is, by the functioning of the apparatus in London, able to

[10] "Echo Suppressors for Long Telephone Circuits". A. B. Clark and R. C. Mathes: *Journal of the American Institute of Electrical Engineers*, Vol. XLIV, June 1925, pp. 618–626. Also "Telephone Repeaters," C. Robinson and R. M. Chamney: *The Electrician*, December 12, 1924, pp. 665–667.

speak through the circuit. Certain interlocking arrangements are provided so that the voice of only one speaker can go through the entire system at one moment. In this way, the two speakers are prevented from talking simultaneously without either one hearing the other. In addition to facilitating two-way operation of the radio channels on the same frequency band the voice operated devices have other valuable features.

The difficulties of reducing the transmission to the narrowest possible band having been overcome, it was necessary to find a free band of this width. Negotiations by the British Post Office people with European stations and by the American Telephone and Telegraph Company with United States stations finally resulted in the moving of a sufficient number of stations to open up a free band having its central frequency at 60 kc. and this frequency is being employed in service.

The above description covers substantially the system which is being used at the present time for giving the commercial transatlantic radio telephone service. This service is not as yet free from difficulties due to unsatisfactory performance of the radio portions of the system. Further development work is being pursued in an attempt to improve these matters. On the English side the British Post Office, after having made comparative measurements of signals and noise in various parts of Scotland, has undertaken and now has under construction a new receiving station at Cupar near Dundee, Scotland. This will provide the greater freedom from radio noise which can be obtained by increasing the latitude of the receiving station. At both the receiving ends of the system, improvements are being made in the directive characteristics of the receiving antennas.

So far very little has been said about the operation of the system. At the New York and London terminals where the transmitting and receiving circuits join, there is a considerable amount of apparatus which includes the automatic switching devices, the repeaters with their gain controls, and a variety of measuring apparatus for determining and maintaining the characteristics of the entire system. This apparatus is under the charge of men called technical operators. Two of them, one in New York and the other in London, have the duty of maintaining the best possible transmission conditions on the system by making the most favorable adjustments. The local transmitting and receiving stations are under their charge in so far as apparatus adjustments which affect the circuit performance are concerned. Communication between the stations and the terminal is provided by means of order wires.

As the circuit passes out of the realm of the technical operator going towards the wire system of the country, it consists of an ordinary two-wire trunk which goes to an operating position in the long distance office. At each end of the circuit two telephone operators are employed. One of these operators makes contact with the telephone network in her country to make ready connections for attachment to the transatlantic link. The other operator directs her attention to the transatlantic link and to dealings with her correspondent at the other end in the way of passing call information, making the final connections, pulling down the connections when subscribers have finished, and so on. From the subscribers' standpoint a call is made in the same way as any other long distance call. He asks for "long distance," gives the information regarding the person he wishes to reach in England and then awaits the return call from the long distance operator. When the person called has been located and the transatlantic link is available, the subscriber receives a ring and is connected with his correspondent. They talk back and forth in exactly the same manner as they would over any wire toll circuit and except for the possibility of occasional noises on the circuit which are obviously of radio origin it is difficult for them to realize that their voices are crossing the Atlantic by radio.

A Study of the Regular Combination of Acoustic Elements, with Applications to Recurrent Acoustic Filters, Tapered Acoustic Filters, and Horns

By W. P. MASON

SYNOPSIS: The use of combinations of tubes to produce interference between sound waves and a suppression of certain frequencies originates with Herschel (1833), and was applied by Quincke to stop tones of definite pitch from reaching the ear. Following the development of electrical filters, G. W. Stewart showed that combinations of tubes and resonators could be devised which would give transmission characteristics at low frequencies similar to electrical filters. The assumptions made by Stewart in the development of his theory are that no wave motion need be considered in the elements, and that the lengths of the elements employed are small compared to the wave-length of sound.

The present paper considers primarily regular combinations of acoustic elements, such as straight tubes, and shows that the equations for recurrent filters, tapered filters and horns can be obtained in this manner. The assumption of no wave motion in the elements, made by Stewart, is removed and also account is taken of the viscosity and heat conduction dissipation. The principal difference between acoustic and electric filters is that the former have an infinite number of bands. The effect of using filters between varying terminal impedances is also determined.

Studying next the combination of filters having the same propagation characteristics but in which the conducting tube areas increase in some regular manner, it is shown that a tapered filter results which has a transforming action in addition to its filtering properties. It is shown that if straight tubes are employed and the distance between successive changes in areas is made small we obtain the horn equations first developed by Webster. The general combination of acoustic elements is then considered, and a proof of several theorems has been given.

STEWART, in a series of papers,[1] has studied the recurrent acoustic filter as an analogue of the electric filter with lumped constants. If due account is taken of the wave motion occurring in the individual elements themselves, it appears that the nearest electrical analogue of the acoustic filter is a combination of electric lines.

In the present paper we study primarily regular combinations of acoustic elements, such as straight tubes, and show that the equations for recurrent filters, tapered filters, and horns can be obtained in this manner. The effect of viscosity and heat conduction dissipation has been taken into account, and a consideration of the effect of varying terminal impedances has been included.

I. Equations of Propagation of a Plane Wave in a Uniform Tube

The propagation of plane waves of sound in uniform tubes has been discussed in a number of places,[2] but generally the results obtained are

[1] *Phys. Rev.*, 20, 528 (1922); 23, 520 (1924); 25, 90 (1925).

[2] Rayleigh's "Theory of Sound," Vol. II, p. 318. Lamb's "The Dynamical Theory of Sound," p. 193.

only a determination of the propagation constant, that is, a determination of the attenuation and phase change per unit length, or as more often stated, the attenuation and velocity characteristics. If we solve the differential equations in the manner first employed by Heaviside in the solution of the equation of the electric line, we obtain one more parameter, namely, the characteristic impedance of the tube.

The differential equation, given by Rayleigh,[2] for the propagation of plane waves of sound in a tube of uniform cross-section is

$$\left(1+\frac{R}{S}\sqrt{\frac{\mu}{2\omega\rho}}\right)\frac{\partial^2\xi}{\partial t^2}+\frac{R}{S}\sqrt{\frac{\mu\omega}{2\rho}}\frac{\partial\xi}{\partial t}=c^2\frac{\partial^2\xi}{\partial x^2}, \tag{1}$$

where ξ denotes the displacement of the fluid at a distance x from one end of the tube,

$\mu =$ the coefficient of viscosity of the medium,

$\rho =$ the density of the medium,

$R =$ perimeter and $S =$ cross-sectional area of pipe,

$\omega =$ frequency of vibration times 2π,

$C = \sqrt{\frac{P_0\gamma}{\rho}} =$ velocity of sound in medium,

$\gamma =$ ratio of specific heats of medium.

This equation is valid for tube diameters and frequencies such that

$$\sqrt{\frac{\rho\omega}{2\mu}}\frac{S}{R} > 1$$

and hence can be used for all frequencies of interest in connection with acoustic filters.

Kirchoff[3] extended the theory to take account of the losses due to heat conduction in the medium. His results indicate that in order to take account of this effect, the square root of the coefficient of viscosity should be replaced by a quantity γ', given by

$$\gamma' = \sqrt{\mu} + \left(\sqrt{\gamma} - \frac{1}{\sqrt{\gamma}}\right)\sqrt{\upsilon},$$

where υ is the coefficient of heat conductivity of the medium. By the kinetic theory of gases υ has the value $5/2\ \mu$.

The most useful solution for our present purpose is obtained by writing

$$\xi = e^{i\omega t}(A\cosh\alpha x + B\sinh\alpha x), \tag{2}$$

[3] Rayleigh, "Theory of Sound," Vol. II, p. 325.

where A and B are constants and α by analogy with an electric line is the propagation constant of the tube. Substituting (2) in (1), we see that (2) is a solution provided

$$\alpha^2 = -\frac{\omega^2}{C^2}\left[\left(1+\frac{R}{S}\sqrt{\frac{\gamma'^2}{2\omega\rho}}\right) - i\frac{R}{S}\sqrt{\frac{\gamma'^2}{2\omega\rho}}\right]. \quad (3)$$

Now α can be written $\alpha = a + ib$, where a is the attenuation constant and b the phase constant. If we solve for a and b, assuming

$$\frac{R}{S}\sqrt{\frac{\gamma'^2}{2\omega\rho}}$$

is a small quantity, we obtain

$$\alpha = a + ib = \frac{1}{2}\frac{R}{CS}\sqrt{\frac{\gamma'^2\omega}{2\rho}} + i\frac{\omega}{C}\left[1 + \frac{1}{2}\frac{R}{S}\sqrt{\frac{\gamma'^2}{2\omega\rho}}\right]. \quad (4)$$

We are generally interested in the volume velocity $S\xi = V$, so we can rewrite equation (2) as

$$V = i\omega S e^{i\omega t}[A \cosh \alpha x + B \sinh \alpha x]. \quad (5)$$

To determine one constant of equation (5), let x equal zero. Then

$$V_{x=0} = V_1 = i\omega e^{i\omega t} SA$$

or

$$A = \frac{V_1}{i\omega S e^{i\omega t}}. \quad (6)$$

We have the additional relation

$$P - P_0 = -P_0\gamma\frac{\partial\xi}{\partial x} = p, \quad (7)$$

where p denotes the excess pressure. Substituting (2) in (7), and differentiating, we have

$$p = -P_0\gamma e^{i\omega t}(A\alpha \sinh \alpha x + B\alpha \cosh \alpha x).$$

Putting $x = 0$, we have

$$p_{x=0} = p_1 = -P_0\gamma e^{i\omega t}(B\alpha)$$

or

$$B = -\frac{p_1}{\alpha P_0 \gamma e^{i\omega t}}. \quad (8)$$

Substituting the value of A and B in (5) and (7), we have

$$V = V_1 \cosh \alpha x - \frac{p_1 i\omega S \sinh \alpha x}{P_0\gamma\alpha}, \qquad p = p_1 \cosh \alpha x - V_1 \frac{(P_0\gamma\alpha)}{i\omega S} \sinh \alpha x. \tag{9}$$

$(P_0\gamma\alpha)/(i\omega)$ is, by analogy with the electric line, the characteristic impedance [4] per square centimeter of the tube. It is the ratio of p_1/ξ_1 for an infinitely long tube. For since $\cosh \alpha x = \frac{1}{2}(e^{\alpha x} + e^{-\alpha x})$ while $\sinh \alpha = \frac{1}{2}(e^{\alpha x} - e^{-\alpha x})$, then when x approaches infinity, and dissipation exists in the tube, $\cosh \alpha x$ approaches $\sinh \alpha x$, and both approach infinity. Hence the ratio of P_1/V_1 equals $P_0\gamma\alpha/i\omega S$. The propagation constant α has the physical significance that $e^{-\alpha x}$ equals the ratio of V to V_1 or p to p_1, when we are dealing with an infinitely long tube, as can be seen by substituting $p_1/V_1 = P_0\gamma\alpha/i\omega S$ in (9) and solving for the above ratios. The real part of α, i.e. a, determines the rate at which the linear or volume velocity, or pressure, decreases with distance, while the imaginary part b determines the phase of pressure or velocity with respect to the initial values, and hence is known as the phase constant and gives the phase rotation per unit length of pipe. Now since the velocity of propagation C' is

$$C' = \frac{\omega}{b},$$

we have by equation (4)

$$C' = C\left[1 - \frac{1}{2}\frac{R}{S}\sqrt{\frac{\gamma'^2}{2\omega\rho}}\right].$$

The attenuation constant and the velocity reduce to the familiar Helmholtz formulæ, for circular sections.[5]

We write (9) as

$$\left.\begin{aligned} V &= V_1 \cosh \alpha x - \frac{p_1 S}{Z_L} \sinh \alpha x, \\ p &= p_1 \cosh \alpha x - \frac{V_1 Z_L}{S} \sinh \alpha x, \end{aligned}\right\} \tag{10}$$

where Z_L represents the specific characteristic impedance $P_0\gamma\alpha/i\omega$.

[4] The analogy between pressure and electromotive force, volume velocity and current, and impedance to ratio of pressure and volume velocity was first pointed out by Webster[9]. Another system in which force and e.m.f., and linear velocity and current are related, is very convenient when we are dealing with combinations of mechanical elements such as masses and elasticities and no account has to be taken of the area. In the first system, the total impedance is Z_L (per sq. cm.) divided by S whereas in the second system it is $Z_L S$. We follow the first system expressing, however, the impedance in terms of the impedance per square centimeter, which is the same on either systems of units.

[5] See Lamb, "Dynamical Theory of Sound," p. 193, or Rayleigh, "Theory of Sound," Vol. II, p. 319.

The value of the specific characteristic impedance $P_0\gamma\alpha/i\omega$ becomes on substituting in the value of α

$$Z_L = \sqrt{P_0\gamma\rho}\left[\left(1+\frac{1}{2}\frac{R}{S}\sqrt{\frac{\gamma'^2}{2\omega\rho}}\right) - i\frac{1}{2}\frac{R}{S}\sqrt{\frac{\gamma'^2}{2\omega\rho}}\right]. \tag{11}$$

If we assume no dissipation, $\gamma' = 0$ and $Z_L = \sqrt{P_0\gamma\rho}$. In any case at fairly high frequencies Z_L approaches $\sqrt{P_0\gamma\rho}$. For example, for air in a circular tube 1 centimeter in diameter, Z_L departs from its final value $\sqrt{P_0\gamma\rho}$ by less than 5 per cent at 100 cycles. The attenuation constant a increases as the square root of the frequency, while the phase constant b is little affected by the dissipation and at high frequencies approaches the value ω/C.

II. Effect of a Junction or of a Change in Area of the Conducting Tube

Suppose that we have a straight conducting tube, with a sidebranch as shown in Fig. 1. Let the excess pressure of the incoming plane wave

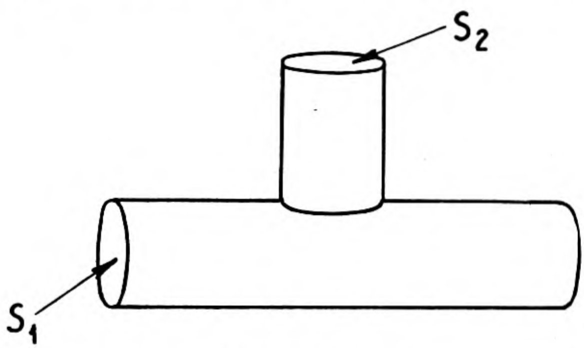


Fig. 1—An acoustic junction

be p_1. The ordinary assumption is that the width of the junction is small compared with a wave-length and hence the pressure is practically constant in the sidebranch, and main branch over the portion in immediate contact with the sidebranch. It states also that the algebraic sum of the volume displacements at a junction of tubes is zero. If S_1 is the area of the main conducting tube, S_2 the area of the branch tube, $\dot{\xi}_1$ the linear velocity of the incoming wave in the conducting tube, $\dot{\xi}_2$ the linear velocity of the outgoing wave from the junction and $\dot{\eta}$ the linear velocity in the branch tube at the junction, we can write the equation

$$\dot{\xi}_1 S_1 = \dot{\xi}_2 S_1 + \dot{\eta} S_2 \qquad \text{or} \qquad V_1 = V_2 + V'.$$

We have now that $\dot{\eta} = p_1/Z_S$ where Z_S is the impedance per unit area of the sidebranch, or the ratio of the excess pressure to the linear velocity. Substituting this value in the above equation, we have

We have also

$$\left.\begin{aligned} V_2 &= V_1 - \frac{p_1 S_2}{Z_S}. \\ p_2 &= p_1, \end{aligned}\right\} \tag{12}$$

where p_2 is the excess pressure in the conducting tube on the outgoing side. The equations are exactly equivalent to Kirchoff's laws, and hence any equation for a combination of acoustic elements will also apply to the combinations of equivalent electric elements.

A slightly better approximation than the above has been obtained by solving completely the case of three pistons placed in the sides of a rectangular box. This corresponds closely to the condition considered here, if we have rectangular tubes, since the waves can be considered plane up to the junction point with little possibility of error. The solution obtained indicates that the main effect of the junction point is to add an end correction to all the tubes entering the junction. For example, we will measure the length of the main conducting tube, between sidebranches, from the center of the sidebranches rather than the edge, as the approximation given first would imply. Also the length of the sidebranch should be measured from the center of the conducting tube, rather than the edge. For other types of junctions, different end corrections will apply to the sidebranch tubes. For example if the width of the junction is large compared to the width of the sidebranch, we should expect Rayleigh's theoretical value of .82 R to apply where R is the radius of the sidebranch tube. Hence the equations for a junction are equivalent to Kirchoff's laws with the additional proviso that end corrections shall be added to tubes entering a junction.

The effect of a change of area of the conducting tube can be obtained with the same assumptions as above. If we have one conducting tube of area S_1, joined to a second of area S_2, we can write

$$\dot{\xi}_1 S_1 = \dot{\xi}_2 S_2 \qquad \text{or} \qquad V_1 = V_2, \tag{13}$$

where $\dot{\xi}_1$ is the linear velocity in the first tube and $\dot{\xi}_2$ in the second tube. We have also that the pressures in the adjoining tubes are equal. Hence

$$p_2 = p_1 \qquad \text{and} \qquad V_2 = V_1. \tag{14}$$

This equation is of the same order of approximation as the second approximation given above for a junction, since we measure the length from one change of area to the next change.

Equation 14 has been found to hold well as long as the change in area is small while equation 12 holds well as long as the length of a junction is less than half of a wave-length.

III. Recurrent Filters

With the aid of equations (10), (12), and (14), we can obtain the propagation characteristics of any structure employing straight tubes, sidebranches, and changes in area of conducting tubes.

Among the simplest of these are recurrent filters. Fig. 2 shows an

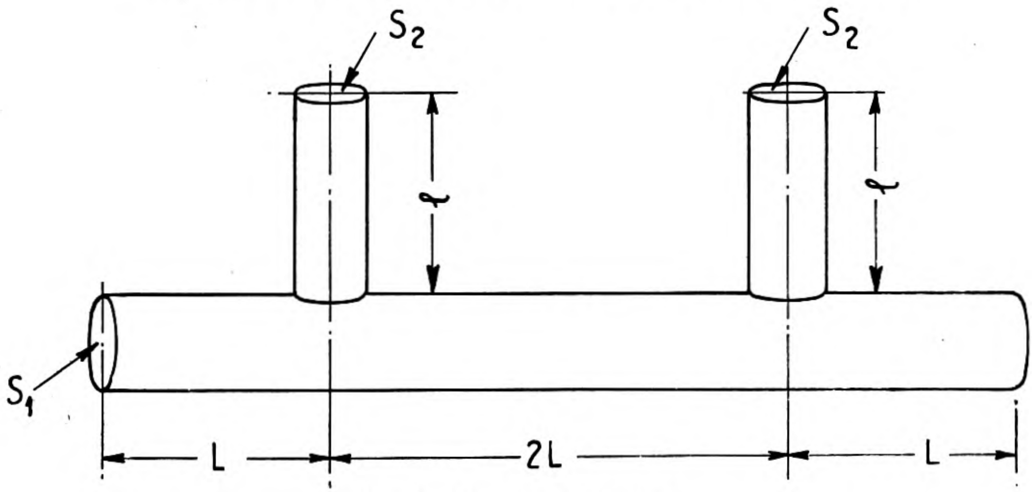


Fig. 2—A typical acoustic filter

example of this type of structure, a main conducting tube, with equally spaced sidebranches. In order to make the structure symmetrical, we let the distance L between one end and the first sidebranch equal one half the distance between two sidebranches. We can then write with regard to the first tube

$$\left.\begin{aligned} V_2 &= V_1 \cosh \alpha_1 L - \frac{p_1}{Z_{L_1}} S_1 \sinh \alpha_1 L, \\ p_2 &= p_1 \cosh \alpha_1 L - V_1 \frac{Z_{L_1}}{S_1} \sinh \alpha_1 L, \end{aligned}\right\} \tag{15}$$

where α_1 and Z_{L_1} refer to the conducting tube. For the junction, we have by (12)

$$\left.\begin{aligned} V_3 &= V_2 - \frac{p_2}{Z_S} S_2, \\ p_3 &= p_2. \end{aligned}\right\} \tag{16}$$

Combining with (15), we have

$$\left.\begin{aligned} V_3 &= V_1\left(\cosh \alpha_1 L + \frac{Z_{L_1}S_2}{Z_S S_1}\sinh \alpha_1 L\right) \\ &\qquad - p_1 S_1\left[\frac{\sinh \alpha_1 L}{Z_{L_1}} + \frac{S_2 \cosh \alpha_1 L}{Z_S S_1}\right], \\ p_3 &= p_1 \cosh \alpha_1 L - \frac{V_1 Z_{L_1}}{S_1}\sinh \alpha_1 L. \end{aligned}\right\} \tag{17}$$

The pressures and volume velocities p_4 and V_4 at one half the distance between the first and second sidebranches are again

$$\left.\begin{aligned} V_4 &= V_3 \cosh \alpha_1 L - \frac{p_3 S_1}{Z_{L_1}}\sinh \alpha_1 L, \\ p_4 &= p_3 \cosh \alpha_1 L - V_3 \frac{Z_{L_1}}{S_1}\sinh \alpha_1 L. \end{aligned}\right\} \tag{18}$$

Combining with (17), we obtain

$$\left.\begin{aligned} V_4 &= V_1\left(\cosh 2\alpha_1 L + \frac{Z_{L_1}S_2}{2Z_S S_1}\sinh 2\alpha_1 L\right) \\ &\qquad - \frac{p_1 S_1}{Z_{L_1}}\left(\sinh 2\alpha_1 L + \frac{Z_{L_1}S_2}{Z_S S_1}\cosh^2 \alpha_1 L\right)\cdot \\ p_4 &= p_1\left(\cosh 2\alpha_1 L + \frac{Z_{L_1}S_2}{2Z_S S_1}\sinh 2\alpha_1 L\right) \\ &\qquad - \frac{V_1 Z_{L_1}}{S_1}\left(\sinh 2\alpha_1 L + \frac{Z_{L_1}S_2}{Z_S S_1}\sinh^2 \alpha_1 L\right)\cdot \end{aligned}\right\} \tag{19}$$

These equations apply to the first section of the filter. By comparison with equation (10) we see that we can write equation (19) as

$$\left.\begin{aligned} V_4 &= V_1 \cosh \Gamma - \frac{p_1 S_1}{Z_0}\sinh \Gamma, \\ p_4 &= p_1 \cosh \Gamma - V_1 \frac{Z_0}{S_1}\sinh \Gamma, \end{aligned}\right\} \tag{20}$$

where

$$\begin{aligned} \cosh \Gamma &= \left(\cosh 2\alpha_1 L + \frac{Z_{L_1}S_2}{2Z_S S_1}\sinh 2\alpha_1 L\right), \\ Z_0 &= Z_{L_1}\sqrt{\frac{1 + \dfrac{Z_{L_1}S_2}{2Z_S S_1}\tanh \alpha_1 L}{1 + \dfrac{Z_{L_1}S_2}{2Z_S S_1}\coth \alpha_1 L}} \end{aligned} \tag{21}$$

and

$$\sinh \Gamma = \sinh 2\alpha_1 L \sqrt{\left(1 + \frac{Z_{L_1} S_2}{2Z_S S_1} \tanh \alpha_1 L\right)\left(1 + \frac{Z_{L_1} S_2}{2Z_S S_1} \coth \alpha_1 L\right)}.$$

Z_0 and Γ are sometimes called the equivalent line parameters. If we have n sections of the type discussed above, we can write n equations of the kind given by (20). If we eliminate all the terms except for the first and last sections, it can be shown that

$$\left.\begin{aligned} V_n &= V_1 \cosh n\Gamma - \frac{p_1 S_1}{Z_0} \sinh n\Gamma, \\ p_n &= p_1 \cosh n\Gamma - \frac{V_1 Z_0}{S_1} \sinh n\Gamma. \end{aligned}\right\} \tag{22}$$

We see then that Γ represents the propagation constant of one section and Z_0 its specific characteristic impedance. They have the physical interpretation, that Z_0 represents the specific impedance looking into an infinite sequence of these sections, while Γ represents the ratio of excess pressure or volume velocity between one section and the next, when we are dealing with an infinite number of sections, or with a finite number, terminated in the characteristic impedance of the filter.

It is customary in electric filter design to determine the characteristics of a dissipationless filter, and to regard dissipation as causing a slight change in the filter characteristic, which usually occurs most prominently in the pass bands. If we neglect dissipation, equation (21) becomes

$$\cosh \Gamma = \left[\cos\left(\frac{2\omega L}{C}\right) + \frac{i\sqrt{P_0 \gamma \rho}\, S_2}{2Z_S S_1} \sin\left(\frac{2\omega L}{C}\right)\right],$$

$$Z_0 = \sqrt{P_0 \gamma \rho} \sqrt{\frac{1 + \dfrac{i\sqrt{P_0 \gamma \rho}\, S_2}{2Z_S S_1} \tan\left(\dfrac{\omega L}{C}\right)}{1 - \dfrac{i\sqrt{P_0 \gamma \rho}\, S_2}{2Z_S S_1} \cot\left(\dfrac{\omega L}{C}\right)}}. \tag{23}$$

The propagation constant Γ is in general a complex number $A + iB$. The real part represents a diminution of the volume velocity or the pressure, while the imaginary part represents a phase change, as can be seen from the fact that the ratio of pressure or volume velocity is

$$\frac{p_2}{p_1} = e^{-\Gamma} = e^{-(A+iB)} = e^{-A} (\cos B - i \sin B).$$

Now $\cosh \Gamma = \cosh (A + iB) = \cosh A \cos B + i \sinh A \sin B$. Hence we see from equation (23), if Z_S is an imaginary quantity, the expression for $\cosh \Gamma$ is always real, and hence either $\sinh A$ or $\sin B$

is always zero. Hence either the attenuation constant A is zero, or the phase shift is zero, π radians or some multiple of π radians. Now since cosh A can never be less than 1 while cos B must lie between $+1$ and -1, then when the expression for cosh Γ is between -1 and $+1$, the attenuation constant A is zero and cos B equals the expression in (23). When the value of cosh Γ is outside the limits ± 1, the phase shift is 0, π, or some multiple and the attenuation constant A is given by the expression in (23).

The specific characteristic impedance Z_0, given in (23), can be shown to be a real quantity within the transmitted band and an imaginary quantity outside the transmitted band.

The type of filter obtained with the structure shown in Fig. 2 depends on the sidebranch impedance Z_S. As long as Z_S is of such a value as to make the expression for cosh Γ greater in magnitude than 1, an attenuation band occurs, while if cosh Γ is less than 1, a pass band occurs. The cut-off frequencies of the band occur when cosh $\Gamma = \pm 1$. From equation (23) the cut-off frequencies occur when

$$Z_S = \frac{i\sqrt{P_0\gamma\rho}S_2}{2S_1}\cot\left(\frac{\omega L}{C}\right) \quad \text{or} \quad Z_S = -\frac{i\sqrt{P_0\gamma\rho}S_2}{2S_1}\tan\left(\frac{\omega L}{C}\right). \qquad (24)$$

A. Low Pass Filter

The model shown in Fig. 2 can be used to obtain the different types of recurrent filters possible by acoustic means. One of the simplest types of filters in the electrical case is the low pass filter. No exact analogue of this filter exists in the acoustic case, as every acoustic filter has more than one band, but a filter which passes low frequencies and attenuates high frequencies can be designed.

Suppose that the sidebranch used is a straight tube closed at one end. Then by equation (10), the impedance Z_S, when the tube is terminated in an infinite impedance, is

$$Z_S = Z_{L_2}\coth \alpha_2 l,$$

where Z_{L_2} and α_2 are respectively the specific characteristic impedance and propagation constant of the sidebranch, and l its length measured to the center of the conducting tube. Substituting this in the expression for cosh Γ and Z_0, we have

$$\left.\begin{aligned} \cosh\Gamma &= \left(\cosh 2\alpha_1 L + \frac{Z_{L_1}S_2 \sinh 2\alpha_1 L}{2Z_{L_2}S_1\coth\alpha_2 l}\right), \\ Z_0 &= Z_{L_1}\sqrt{\frac{1+\dfrac{Z_{L_1}S_2\tanh\alpha_1 L}{2Z_{L_2}S_1\coth\alpha_2 l}}{1+\dfrac{Z_{L_1}S_2\coth\alpha_1 L}{2Z_{L_2}S_1\coth\alpha_2 l}}}. \end{aligned}\right\} \qquad (25)$$

If we assume no dissipation, and substitute the values of α_1 and Z_L given in section (I), we have

$$\cosh \Gamma = \left[\cos\left(\frac{2\omega}{C} L\right) - \frac{S_2 \sin\left(\frac{2\omega}{C} L\right)}{2S_1 \cot\left(\frac{\omega}{C} l\right)} \right], \tag{26}$$

$$Z_0 = \sqrt{P_0 \gamma \rho}\; \frac{\sqrt{1 - \frac{S_2}{2S_1}\left(\frac{\tan \frac{\omega}{C} L}{\cot \frac{\omega}{C} l}\right)}}{\sqrt{1 + \frac{S_2}{2S_1}\left(\frac{\cot \frac{\omega}{C} L}{\cot \frac{\omega}{C} l}\right)}}. \tag{27}$$

An example of the type of filter obtained by acoustic means, is given when we let $l = 3L$. Fig. 3 gives a plot of the value of Γ for several ratios of S_2/S_1. Fig. 4 shows the corresponding values of the specific characteristic impedance Z_0.

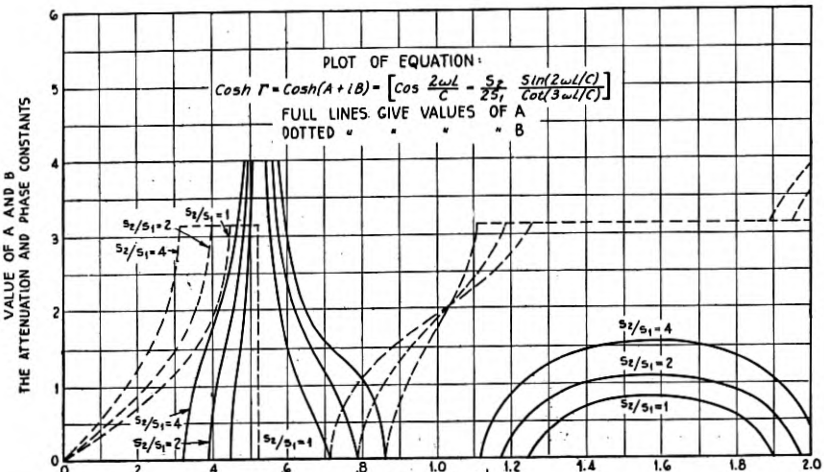


Fig. 3—Propagation constants for a low pass type of filter

A knowledge of Γ will determine the ratio of pressures or volume velocities, if we have an infinite sequence of sections, or if we terminate a finite sequence in the impedance Z_0. If however the terminating impedance is not the characteristic impedance, $e^{-\Gamma}$ no longer represents the ratios of pressures between adjacent sections.

What is generally desired is a knowledge of the effect produced by inserting the filter in a given acoustic system. With the aid of Thévenin's theorem, which is proved for an acoustic system in Appendix I, and equations (20) and (21), this effect can be obtained. Thévenin's

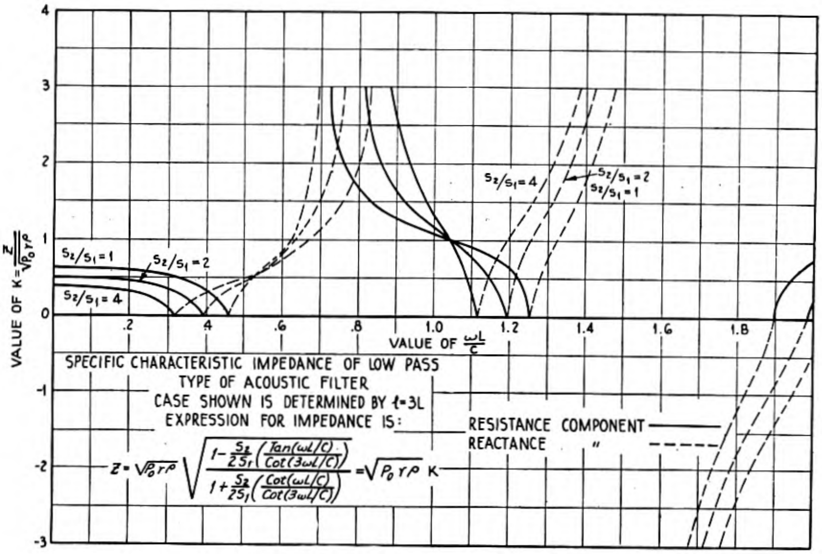


Fig. 4—Specific characteristic impedances for a low pass type of filter

theorem states: If a source of simple harmonic pressure p_0 and of internal impedance Z_T, per square centimeter, is connected to an acoustic system, and if the specific impedance Z_R terminates the system, the volume velocity at the termination of the system will be $p_0'/[(Z_T'/S_1) + (Z_R/S_n)]$, where p_0' is the pressure at the terminating end when this is closed through an infinite impedance, and Z_T' is the impedance per sq. cm. looking back into the acoustic system when this terminated in the impedance Z_T. S_1, and S_n are the areas at the input and output junctions, respectively.

Making use of Thévenin's theorem, the effect of inserting a filter in a given system is the same as the effect obtained by inserting this filter between a source of pressure p_0, with an internal impedance of Z_a/S_1 and a terminating impedance Z_b/S_n, where Z_a/S_1 and Z_b/S_n are respectively the total impedances looking toward the source, and away from the source at the insertion junction of the acoustic system. We have from equation (20)

$$V_2 = V_1 \cosh \Gamma - \frac{p_1 S_1}{Z_0} \sinh \Gamma.$$

$$p_2 = p_1 \cosh \Gamma - \frac{V_1 Z_0}{S_1} \sinh \Gamma,$$

Making use of the above, we can write

$$p_0 = p_1 + \frac{V_1 Z_a}{S_1}.$$

Substituting this, the above equation takes the form

$$\left.\begin{aligned} p_2 &= p_0 \cosh \Gamma - \frac{V_1}{S_1}[Z_0 \sinh \Gamma + Z_a \cosh \Gamma], \\ V_2 &= V_1\left(\cosh \Gamma + \frac{Z_a}{Z_0}\sinh \Gamma\right) - \frac{P_0 S_1}{Z_0}\sinh \Gamma. \end{aligned}\right\} \tag{28}$$

Eliminating V_1 and substituting $V_2 Z_b/S_1$ for p_2, since here the area remains constant at the two junctions, we have

$$V_2 = \frac{p_0 S_1}{\left[Z_b \cosh \Gamma + \dfrac{Z_a Z_b}{Z_0}\sinh \Gamma + Z_a \cosh \Gamma + Z_0 \sinh \Gamma\right]}.$$

The most useful way of writing this equation is

$$V_2 = \left(\frac{p_0 S_1}{2Z_b}\right)\left(\frac{2Z_0}{Z_0 + Z_a}\right)\left(\frac{2Z_b}{Z_0 + Z_b}\right)(e^{-\Gamma}) \times \left[\frac{1}{1 - e^{-2\Gamma}\left(\dfrac{Z_0 - Z_a}{Z_0 + Z_a}\right)\left(\dfrac{Z_0 - Z_b}{Z_0 + Z_b}\right)}\right]. \tag{29}$$

The volume velocity in the termination of the acoustic system, if the filter were not inserted, is obviously $p_0/[(Z_a/S_1) + (Z_b/S_1)]$. Hence the effect of inserting the filter at any junction is to change the volume velocity of the system by the factor

$$\left(\frac{Z_a + Z_b}{2Z_b}\right)\left(\frac{2Z_0}{Z_0 + Z_a}\right)\left(\frac{2Z_b}{Z_0 + Z_b}\right)(e^{-\Gamma}) \times \left[\frac{1}{1 - e^{-2\Gamma}\left(\dfrac{Z_0 - Z_a}{Z_0 + Z_a}\right)\left(\dfrac{Z_0 - Z_b}{Z_0 + Z_b}\right)}\right]. \tag{30}$$

A physical interpretation of equation (30) can be obtained in terms of the transmission and reflection factors first introduced by Heaviside.[6] Heaviside showed that at a junction, a reflection of a wave takes place if the impedances looking towards the source and away from the source are not equal. He showed that the current reflected on striking a junction, will be the unmodified current in the line multiplied by the

[6] Heaviside "Electromagnetic Theory" Vol. II, page 79.

factor, $(Z_I - Z_T)/(Z_I + Z_T)$, while the current transmitted to the terminating side of the junction will be the unmodified current in the line multiplied by the factor $2Z_T/(Z_I + Z_T)$ where Z_I and Z_T are respectively the impedances looking towards and away from the source at the junction. We see then that the second and third factors are transmission factors, determining respectively the transmission from the input impedance Z_a to the inserted structure, and from the inserted structure to the output impedance Z_b. The first factor is the inverse of the transmission factor determining the transmission from the impedance Z_a to the impedance Z_b. The fourth factor is the transfer factor and gives the reduction in volume velocity due to attenuation. The fifth factor has been called the interaction factor, and it gives the change in volume velocity in the termination due to repeated reflections of the volume velocity within the structure. All of these factors reduce to 1 except the transfer factor when $Z_a = Z_b = Z_0$. It will be noted that all factors except the transfer factor cancel out if $Z_a = Z_0$, or $Z_b = Z_0$.

The effect on the pressure due to inserting a filter can be shown to be given also by equation (30).

If the terminating impedances are resistances about equal to an average of the resistance value of Z_0, the effect of these is generally to introduce some loss in the pass band, when the characteristic impedance differs materially from the terminating impedances due to a reflection of the sound wave at the junction points. Since the characteristic impedance of a non-dissipative filter goes either to zero or infinity at the cut-off frequency, the effect of the reflection loss is generally to narrow the pass bands of the filter.

The effect of dissipation, when we take account of the viscosity effects by equations (20) or (21), is two-fold. It changes slightly the position of the band in the frequency range, due to a small change in the velocity of propagation. This is generally negligible. The other effect is to introduce attenuation in the pass band, due to absorption and dissipation of the sound wave.

B. High Pass Filter

An analogous type of high pass filter, which will attenuate the low frequencies and pass the high frequencies, can be made from the structure shown in Fig. 2 by using side tubes which are open on the outer end. The termination at the end of an open tube has been shown by Rayleigh [7] to be a mass with some resistance due to radiation. We could substitute this relation in equation (10) to determine

[7] Rayleigh, "Theory of Sound," Vol. II, p. 106.

the impedance Z_S looking into the sidebranch. Another approximation used with organ pipes is to consider the tube extended by a length .57 times the radius of the tube, and to consider this extended tube terminated in a zero impedance.

The impedance Z_S for this case is from (10)

$$\frac{p_1 S_2}{V_1} = Z_S = Z_{L_2} \tanh \alpha_2 l',$$

where l' is the corrected length of the pipe. Substituting this value in equation (21), we have

$$\cosh \Gamma = \left[\cosh 2\alpha_1 L + \frac{Z_{L_1} S_2 \sinh 2\alpha_1 L}{2 Z_{L_2} S_1 \tanh \alpha_2 l'} \right],$$

$$Z_0 = Z_{L_1} \sqrt{\frac{1 + \dfrac{Z_{L_1} S_2 \tanh \alpha_1 L}{2 Z_{L_2} S_1 \tanh \alpha_2 l'}}{1 + \dfrac{Z_{L_1} S_2 \coth \alpha_1 L}{2 Z_{L_2} S_1 \tanh \alpha_2 l'}}}. \tag{31}$$

For no dissipation these equations become

$$\cosh \Gamma = \left[\cos\left(\frac{2\omega L}{C}\right) + \frac{S_2}{2S_1} \left[\frac{\sin\left(\dfrac{2\omega L}{C}\right)}{\tan\left(\dfrac{\omega l'}{C}\right)} \right] \right],$$

$$Z_0 = \sqrt{P_0 \gamma \rho} \sqrt{\frac{1 + \dfrac{S_2 \tan\left(\dfrac{\omega L}{C}\right)}{2S_1 \tan\left(\dfrac{\omega l'}{C}\right)}}{1 - \dfrac{S_2 \cot\left(\dfrac{\omega L}{C}\right)}{2S_1 \tan\left(\dfrac{\omega l'}{C}\right)}}}.$$

Fig. 5 shows a plot of Γ for several ratios of S_2/S_1, when $l' = 3L$.

C. Band Pass Type of Filter

The high pass type of filter discussed above can also be considered as a band pass type of filter, in that an attenuation occurs at zero frequency, then a pass band, and a second attenuation band. A different arrangement of the pass bands can be obtained from the structure shown in Fig. 2, by inserting two sidebranches at one junction point, one of which is open at the outside end and the other closed.

An example of the type of characteristic obtained, is given by the special case where the lengths of both tubes are the same and equal to

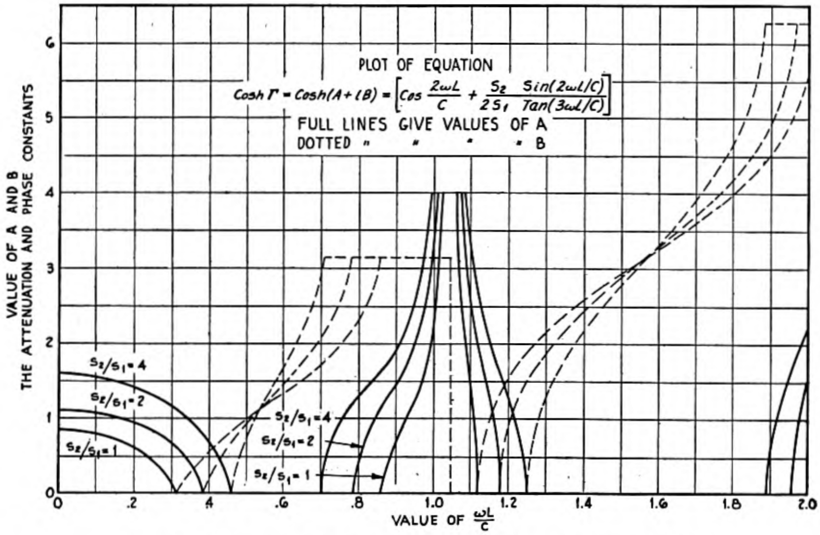


Fig. 5—Propagation constants for a high pass type of filter

$3L$. If S_2 is the area of the open tube and S_3 that of the closed tube, then neglecting dissipation, we find

$$\cosh \Gamma = \cos \frac{2\omega L}{C} + \frac{1}{2S_1}\left[\frac{S_2}{\tan \frac{3\omega L}{C}} - \frac{S_3}{\cot \frac{3\omega L}{C}}\right] \sin \frac{2\omega L}{C}.$$

A plot of A, the attenuation constant, for several values of S_2/S_1 and S_3/S_1 is given in Fig. 6.

D. *Other Types of Sidebranches*

We have so far considered only the characteristics obtained where we employ straight tubes. A number of cases can be solved in which the elements employed are not straight tubes although we cannot take account of the viscosity dissipation in these cases. As an example, the characteristics of a filter will be worked out, which employs a straight tube for the conducting tube and conical tubes closed on the end for the sidebranches. We can make use of equation (21) to determine Γ and Z_0, if we insert the proper value of Z_S for the conical tube.

It is evident that for a conical tube, the proper type of wave is a spherical wave, in place of the plane wave employed for a straight tube. For this case we can write [8] for a simple harmonic wave

$$\frac{\partial^2(r\varphi)}{\partial t^2} = C^2\frac{\partial^2(r\varphi)}{\partial r^2}\,; \quad \dot{\eta} = -\frac{\partial\varphi}{\partial r} \text{ and } \frac{p}{\rho_0} = \dot{\varphi},$$

where φ is the velocity potential, $\dot{\eta}$ the linear velocity for the spherical wave, p the pressure, ρ the average density of the medium, and r the

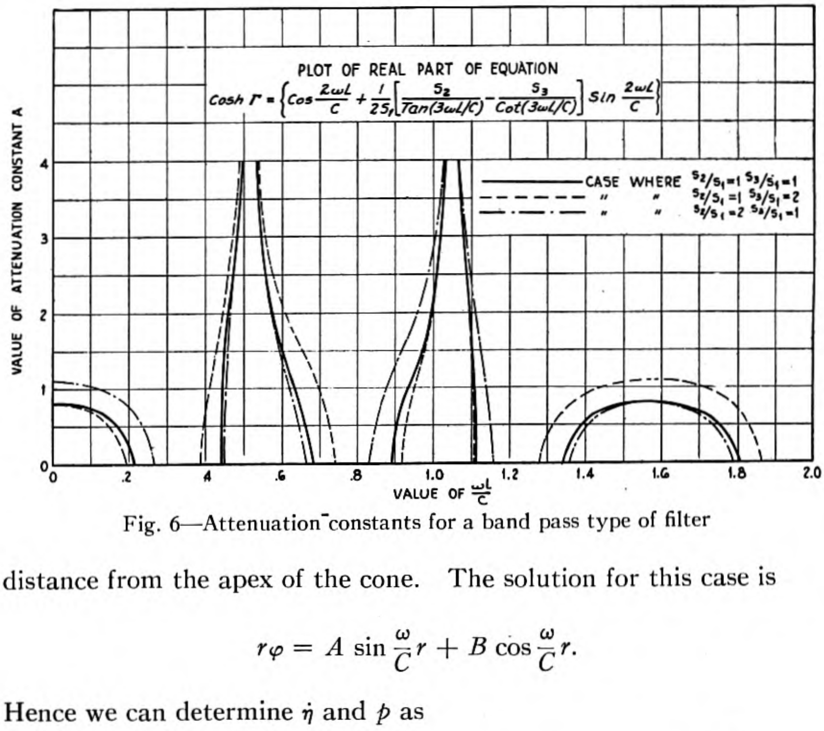


Fig. 6—Attenuation constants for a band pass type of filter

distance from the apex of the cone. The solution for this case is

$$r\varphi = A\sin\frac{\omega}{C}r + B\cos\frac{\omega}{C}r.$$

Hence we can determine $\dot{\eta}$ and p as

$$\dot{\eta} = A\left[\frac{\sin\frac{\omega}{C}r}{r^2} - \frac{\frac{\omega}{C}\cos\frac{\omega}{C}r}{r}\right] + B\left[\frac{\cos\frac{\omega}{C}r}{r^2} + \frac{\frac{\omega}{C}\sin\frac{\omega}{C}r}{r}\right]$$

and

$$p = i\omega\rho\left[\frac{A\sin\frac{\omega}{C}r + B\cos\frac{\omega}{C}r}{r}\right].$$

If now we set $\dot{\eta} = 0$ when $r = x_2$ and determine the ratio of $p/\dot{\eta}$ at

[8] Lamb, "Dynamical Theory of Sound," p. 206. Rayleigh, "Theory of Sound," Vol. II, p. 114.

$r = x_1$, we find

$$Z_S = \frac{p}{\dot{\eta}}$$

$$= -i\sqrt{P_0\gamma\rho}\left[\frac{\cos\frac{\omega}{C}(x_2 - x_1) - \dfrac{\sin\frac{\omega}{C}(x_2 - x_1)}{\frac{\omega}{C}x_2}}{\cos\frac{\omega}{C}(x_2 - x_1)\left[\dfrac{1}{\frac{\omega}{C}x_2} - \dfrac{1}{\frac{\omega}{C}x_1}\right] + \sin\frac{\omega}{C}(x_2 - x_1)\left[1 + \dfrac{1}{\frac{\omega^2}{C^2}x_1x_2}\right]}\right]. \quad (32)$$

If we substitute this value of Z_S in equation (21), we can readily determine the value of Γ and Z_0. Fig. 7 shows a plot of A and B for this case assuming $(x_2 - x_1) = L$.

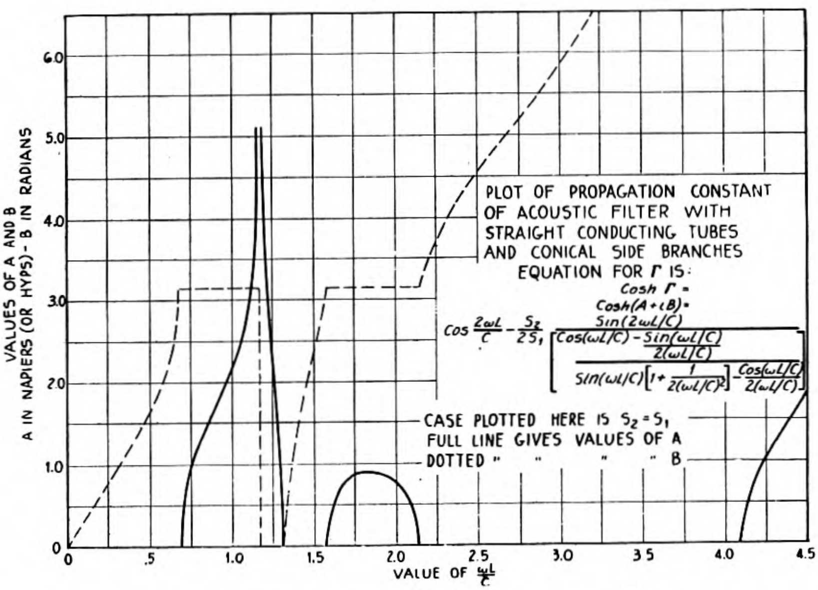


Fig. 7—Propagation constant of a low pass type of filter.

IV. Tapered Filter Structures and Horns

In addition to recurrent filters, other types of filters exist. If, for example, we connect sections with the same propagation constants and characteristic impedances, but whose conducting tube areas increase in some regular manner, a tapered filter is obtained whose characteristics differ from those of a recurrent filter. The distinguishing property introduced by a tapered filter, in addition to its filtering property, is a transformer action which increases the pressure by a given ratio and decreases the volume velocity in the same ratio, or vice versa, thus giving a transforming action and a complete transmission of power over the pass band. This is a useful property, if acoustic systems of different impedances are to be connected together. Horns are the limiting cases of tapered acoustic filters and hence their study has considerable practical importance.

The typical section of a tapered filter considered here is one built up from two symmetrical structures with the same propagation constants and characteristic impedances per square centimeter, but with different cross-sectional areas. If we use any of the recurrent filters discussed in Section III, then, for example, since

$$\cosh \Gamma = \left[\cos\left(\frac{2\omega L}{C}\right) - \frac{S_2 \sin\left(\frac{2\omega L}{C}\right)}{2S_1 \cot \frac{\omega l}{C}} \right]$$

for the low pass filter, to keep the same value of Γ when we vary the conducting tube area it will be necessary to keep the ratio of the areas constant and to leave all values of L and l the same. Similarly for the other types of filters.

If $\Gamma/2$ is the propagation constant of each of the symmetrical structures, Z_0 the characteristic impedance per square centimeter for each structure, S_1 the cross-sectional area of the first structure and S_2 that of the second, we can write three sets of equations for the two structures and the junction point. These are

$$\left.\begin{aligned} p_1' &= p_1 \cosh\frac{\Gamma}{2} - V_1 \frac{Z_0}{S_1} \sinh\frac{\Gamma}{2}, \\ V_1' &= V_1 \cosh\frac{\Gamma}{2} - \frac{p_1 S_1}{Z_0} \sinh\frac{\Gamma}{2}, \end{aligned}\right\}$$

$$p_1'' = p_1'; \quad V_1'' = V_1',$$

$$\left.\begin{aligned} p_2 &= p_1'' \cosh\frac{\Gamma}{2} - V_1'' \frac{Z_0}{S_2} \sinh\frac{\Gamma}{2}, \\ V_2 &= V_1'' \cosh\frac{\Gamma}{2} - \frac{p_1'' S_2}{Z_0} \sinh\frac{\Gamma}{2}. \end{aligned}\right\}$$

Combining these equations, we obtain

$$\left.\begin{aligned} p_2 &= p_1\left[\left(\frac{S_1+S_2}{2S_2}\right)\cosh\Gamma + \left(\frac{S_2-S_1}{2S_2}\right)\right] \\ &\qquad\qquad - \frac{V_1Z_0}{S_1}\left(\frac{S_1+S_2}{2S_2}\right)\sinh\Gamma, \\ V_2 &= V_1\left[\left(\frac{S_1+S_2}{2S_1}\right)\cosh\Gamma - \left(\frac{S_2-S_1}{2S_1}\right)\right] \\ &\qquad\qquad - \frac{p_1S_1}{Z_0}\left(\frac{S_1+S_2}{2S_1}\right)\sinh\Gamma, \end{aligned}\right\} \tag{33}$$

or for simplicity we write

$$\left.\begin{aligned} p_2 &= p_1A - \frac{V_1Z_0}{S_1}B, \\ V_2 &= V_1C - \frac{p_1S_1}{Z_0}D. \end{aligned}\right\} \tag{34}$$

In order to express the propagation in terms of some known functions we will first obtain some relations between the impedances of the sections and the ratios of p_2/p_1 and V_2/V_1. We can write the above equations as

$$\frac{p_2}{p_1} = A - \frac{Z_0}{Z_1}B, \quad \frac{V_2}{V_1} = C - \frac{Z_1}{Z_0}D,$$

where $Z_1/S_1 = p_1/V_1$. Eliminating Z_1, we have

$$\frac{p_2}{p_1}\frac{V_2}{V_1} - A\frac{V_2}{V_1} - C\frac{p_2}{p_1} = BD - AC = -1 \tag{35}$$

as can be seen by multiplying together the above expressions. Solving for the ratio of V_2/V_1 in terms of p_2/p_1, A and C, we have

$$\frac{V_2}{V_1} = \frac{C\dfrac{p_2}{p_1} - 1}{\dfrac{p_2}{p_1} - A}.$$

Multiplying both sides by p_1/p_2, we have

$$\frac{p_1}{V_1}\frac{V_2}{p_2} = \frac{C - \dfrac{p_1}{p_2}}{\dfrac{p_2}{p_1} - A}.$$

Now since $\frac{p_1}{V_1} = \frac{Z_1}{S_1}$ and $\frac{p_2}{V_2} = \frac{Z_2}{S_2}$, we have

$$\frac{Z_2}{S_2} = \frac{Z_1}{S_1}\left[\frac{\frac{p_2}{p_1} - A}{C - \frac{p_1}{p_2}}\right]. \tag{36}$$

Z_2/S_2 and Z_1/S_1 then are respectively the terminating impedance and input impedance necessary to give a structure specified by the factors A, B, C, D the pressure ratio p_2/p_1. To solve for the input impedance we take the first of equations (34) and obtain

$$Z_1 = \frac{Z_0 B}{A - \frac{p_2}{p_1}}. \tag{37}$$

Hence by virtue of (36), the terminating impedance Z_2/S_2 becomes

$$\frac{Z_2}{S_2} = \frac{\frac{Z_0}{S_1} B}{\frac{p_1}{p_2} - C}. \tag{38}$$

Equations (37) and (38) state that there is a relation between the input impedance and the pressure ratio, and the output impedance and the pressure ratio. When one is specified and the constants of the section Z_0, A, B, C, D are given, the others are known.

Suppose now that we wish to join a second structure of this type to the first, assuming that the cross-sectional area at the junction is the same for both. We must have now Z_2, the specific output impedance of the first section, equal to Z_1', the specific input impedance of the second section. Hence we can write.

$$\frac{\frac{Z_0}{S_1} B}{\frac{p_1}{p_2} - C} = \frac{\frac{Z_0}{S_2} B'}{\left(A' - \frac{p_3}{p_2}\right)} \text{ or } \frac{S_2}{S_1} B\left[A' - \frac{p_3}{p_2}\right] = B'\left[\frac{p_1}{p_2} - C\right],$$

where the primes refer to the constants of the second section and where p_3/p_2 is the pressure ratio of the second section. Substituting in the values of A', B, B', C, we have

$$\left(\frac{(S_1 + S_2)(S_2 + S_3)}{2S_1S_3}\right)\cosh\Gamma + \frac{S_1S_3 - S_2^2}{2S_1S_3} = \left(\frac{S_2 + S_3}{2S_3}\right)\frac{p_1}{p_2} + \left(\frac{S_1 + S_2}{2S_1}\right)\frac{p_3}{p_2}. \tag{39}$$

Equation (39) gives the relationship between p_1/p_2 and p_3/p_2 which must be satisfied if the output impedance of one section equals the input impedance of the next section. If we specify a value of p_2/p_1, then the value of p_3/p_2 is determined. The impedance Z_2' terminating the second section is also determined and hence the pressure ratio of the third section, etc. Hence if we specify a value of p_2/p_1, we also determine the propagation characteristic of any other section in a series of sections. The pressure ratios will not in general be constant from section to section.

We can write $p_2/p_1 = Ke^{-\delta}$ since this will represent any phase or amplitude change. Similarly we can write p_3/p_2 as $K'e^{-\delta'}$. Substituting these values in (39), we have

$$\left(\frac{(S_1 + S_2)(S_2 + S_3)}{2S_1S_3}\right) \cosh \Gamma + \frac{S_1S_3 - S_2^2}{2S_1S_3} = \left(\frac{S_2 + S_3}{2S_3}\right)\frac{e^{\delta}}{K} + \left(\frac{S_1 + S_2}{2S_1}\right) K'e^{-\delta'}. \tag{40}$$

Now if the value of δ remains unchanged from section to section a great simplification results, for in order to determine the overall pressure ratio we have only to multiply the number of sections by δ. Hence it is desirable to determine for what rate of taper this condition is met and also how good an approximation it is for all rates of taper.

If we set $\delta = \delta'$ and multiply through by $e^{-\delta}$, we obtain

$$e^{-2\delta} - \left[\frac{\left(\frac{(S_1 + S_2)(S_2 + S_3)}{2S_1S_3}\right) \cosh \Gamma + \frac{S_1S_3 - S_2^2}{2S_1S_3}}{\frac{S_1 + S_2}{2S_1} K'}\right] e^{-\delta} + \frac{\left(\frac{S_2 + S_3}{2S_3}\right)\frac{1}{K}}{\left(\frac{S_1 + S_2}{2S_1}\right) K'} = 0.$$

Similarly the equation for the next two sections is

$$e^{-2\delta''} - \frac{\left[\left(\frac{(S_2 + S_3)(S_3 + S_4)}{2S_2S_4}\right) \cosh \Gamma + \left(\frac{S_2S_4 - S_3^2}{2S_2S_4}\right)\right]}{\left(\frac{S_2 + S_3}{2S_2}\right) K''} e^{-\delta''} + \frac{\left(\frac{S_3 + S_4}{2S_4}\right)\frac{1}{K'}}{\left(\frac{S_2 + S_3}{2S_2}\right) K''} = 0.$$

If we are to have $\delta = \delta''$, we must have

$$e^{-\delta}\left[\left[\frac{S_2+S_3}{S_3K'}-\frac{S_3+S_4}{S_4K''}\right]\cosh\Gamma + \left[\frac{S_1S_3-S_2^2}{(S_1+S_2)S_3K'}-\frac{S_2S_4-S_3^2}{(S_2+S_3)S_4K''}\right]\right] = \left[\frac{(S_2+S_3)S_1}{S_3(S_1+S_2)KK'}-\frac{(S_3+S_4)S_2}{(S_2+S_3)S_4K'K''}\right]. \quad (41)$$

Since the term on the left is complex, while that on the right is a numeric, each must separately vanish if we are to have this equality. Similarly the terms within the bracket of the left hand side would have to vanish.

We see that the two terms on the left do not vanish simultaneously unless we satisfy the progression equation

$$2S_1S_3^2 - 2S_2^2S_4 + (S_3 - S_2)(S_1S_4 + S_2S_3) = 0. \quad (42)$$

This equation is satisfied by a system whose area increases exponentially with the distance. The terms involving $S_1S_3 - S_2^2$ and $S_2S_4 - S_3^2$ are always very small no matter what the rate of progression. Hence is is desirable to see if neglecting these terms we can still satisfy the above conditions. The most useful value of the two terms on the right hand side of equation (41) is 1. Hence setting each term equal to 1 and solving for K' and K'', we find that

$$K' = \sqrt{\frac{S_2}{S_3}}; \quad K'' = \sqrt{\frac{S_3}{S_4}}.$$

We see then that if we neglect second order quantities, we can represent with good approximation the pressure ratio of any tapered filter by the expression

$$\frac{p_2}{p_1} = \sqrt{\frac{S_n}{S_{n+1}}}\,e^{-\delta},$$

where δ is the propagation constant of a tapered structure. For a complete solution, δ is not constant except for a progression which satisfies equation (42).

A. Exponentially Tapered Filters and Horns

If we assume that the area of a given section is e^{2t} times as large as that of the section preceding, equation (40) reduces to

$$2e^{-t}[\cosh\Gamma\cosh t] = K'e^{-\delta} + \left(e^{-2t}\times\frac{1}{K}\right)e^{\delta}. \quad (43)$$

We choose now $K' = \sqrt{\frac{S_2}{S_3}} = e^{-t}$ and $K = \sqrt{\frac{S_1}{S_2}} = e^{-t}$.

Then

$$\cosh \delta = \cosh \Gamma \cosh t.$$

To show that δ is the propagation constant for an infinite sequence of such sections, it is necessary to show that δ is the same for any two sections. But equation (43) holds good for any two sections, and hence δ is the same, and represents a solution for an infinite sequence of sections. Now

$$e^{-\delta} = \cosh \delta - \sinh \delta = \cosh \Gamma \cosh t - \sqrt{\sinh^2 \Gamma \cosh^2 t + \sinh^2 t}.$$

Hence

$$\frac{p_2}{p_1} = Ke^{-\delta} = e^{-t}[\cosh \Gamma \cosh t - \sqrt{\sinh^2 \Gamma \cosh^2 t + \sinh^2 t}]$$

and

$$\frac{V_2}{V_1} = e^{-t}e^{-\delta}\left[\frac{C - e^t e^\delta}{e^{-t}e^{-\delta} - A}\right]$$
$$= e^t[\cosh \Gamma \cosh t - \sqrt{\sinh^2 \Gamma \cosh^2 t + \sinh^2 t}],$$

and hence the pressure and volume velocity have the same propagation constant δ but an inverse multiplying factor.

The specific impedance Z_1, looking into a given section, is by equation (37)

$$Z_1 = \frac{Z_0 B}{A - Ke^{-\delta}} = Z_0\left[\frac{\sqrt{\tanh^2 t + \sinh^2 \Gamma} - \tanh t}{\sinh \Gamma}\right] \qquad (44)$$

and similarly Z_2, the specific terminating impedance, can be shown equal to Z_1. Hence the impedance per square centimeter at the junction points is the same for each section.

To observe the action of a tapered filter, let us obtain the product of the pressure by the volume velocity and see how these are propagated. Since the specific impedance is the same from section to section, this will represent also the power propagation. Now since

$$\cosh \delta = \cosh \Gamma \cosh t,$$

a pass band occurs when $1 \geqq \cosh \delta \geqq -1$ and hence the band occurs only when Γ is imaginary, since $\cosh \Gamma < 1$ and > -1, or when the filter repeated recurrently is in its pass band. Furthermore the pass band for the tapered structure will not be as wide as that for a similar recurrent structure, since for the tapered structure the band occurs when $\cosh \Gamma = \pm 1/\cosh t$ while in the recurrent structure, the band occurs when $\cosh \Gamma = \pm 1$. One result of this is that no low pass filter exists in exponentially tapered structures.

Considering now the pressure and volume velocity ratios when $1 \geqq \cosh \delta \geqq -1$, the absolute value of $e^{-\delta}$ is 1. Hence over the band the ratios of pressure and of volume velocity from section to section are respectively e^{-t} and e^{t} or $\sqrt{S_1/S_2}$ and $\sqrt{S_2/S_1}$. Hence one section multiplies the pressure by a ratio $\sqrt{S_1/S_2}$, and the volume velocity by the factor $\sqrt{S_2/S_1}$. Therefore a tapered structure of this kind is equivalent to a transformer of turns ratio $\sqrt{S_1/S_2}$, and a filter of somewhat narrower bands than for the filter repeated recurrently.

To specify completely a filter of this type requires three parameters. Two such parameters have been developed above and are δ, the propagation constant of a tapered filter, and Z_{R_1}, the specific recurrent impedance in one direction. These are given by

$$\left.\begin{aligned} \cosh \delta &= \cosh \Gamma \cosh t, \\ Z_{R_1} &= Z_0 \left[\frac{\sqrt{\tanh^2 t + \sinh^2 \Gamma} - \tanh t}{\sinh \Gamma} \right]. \end{aligned}\right\} \tag{45}$$

We take as the third parameter Z_{R_2}, the specific recurrent impedance in the opposite direction. We can readily determine that Z_{R_2}, the impedance looking in the opposite direction from that used to specify Z_{R_1}, but obtained at the same junction point, is

$$Z_{R_2} = \frac{Z_0 B}{\dfrac{p_1'}{p_2'} - A}.$$

It is desirable to have the same propagation constant serve for the two directions, hence we let $p_2'/p_1' = K_1 e^{-\delta}$. Since K represents a transformer change of the pressure in one direction, we find, when going in the opposite direction, that the pressure should change by the inverse of K, so $K_1 = 1/\text{K}$. Substituting these values for p_1'/p_2',

$$Z_{R_2} = \frac{Z_0 B}{K e^{\delta} - A}. \tag{46}$$

Hence for an exponentially tapered filter

$$Z_{R_2} = Z_0 \left[\frac{\sqrt{\tanh^2 t + \sinh^2 \Gamma} + \tanh t}{\sinh \Gamma} \right]. \tag{47}$$

In terms of the parameters, δ, Z_{R_1} and Z_{R_2} we can express p_2, p_1, V_2, and V_1 as

$$
\begin{aligned}
p_2 &= e^{-t}\left[p_1\left[\cosh\delta + \left[\frac{Z_{R_2} - Z_{R_1}}{Z_{R_2} + Z_{R_1}}\right]\sinh\delta\right]\right. \\
&\qquad \left. - \frac{V_1}{S_1}\left[\frac{2Z_{R_1}Z_{R_2}}{Z_{R_1} + Z_{R_2}}\right]\sinh\delta\right], \\
V_2 &= e^{t}\left[V_1\left[\cosh\delta + \left[\frac{Z_{R_1} - Z_{R_2}}{Z_{R_1} + Z_{R_2}}\right]\sinh\delta\right]\right. \\
&\qquad \left. - p_1S_1\left[\frac{2\sinh\delta}{Z_{R_1} + Z_{R_2}}\right]\right].
\end{aligned} \tag{48}
$$

If now the elements of our structure are non-dissipative straight tubes, instead of a general filter structure, and the length of these tubes between changes of area is made very small, it is evident that the structure reduces to an exponential horn. We now let the ratio $S_1/S_2 = e^{-2t}$ be expressed as

$$\frac{S_1}{S_2} = e^{-2Tl} = e^{-2t},$$

where l is the distance between changes in area and T a new taper constant. Then Γ, for a straight tube, neglecting dissipation, becomes $\Gamma = i\omega l/c$ and hence

$$
\begin{aligned}
\cosh\delta &= \cosh\frac{i\omega l}{C}\cosh Tl \\
&= \left(1 - \frac{\omega^2l^2}{2!C^2} + \frac{\omega^2l^2}{4!C^4} + \cdots\right)\left(1 + \frac{(Tl)^2}{2!} + \frac{(Tl)^4}{4!} + \cdots\right) \\
&= 1 + \frac{l^2\left(T^2 - \frac{\omega^2}{C^2}\right)}{2!} + l^4\frac{\left(T^4 - 6\frac{\omega^2}{C^2}T^2 + \frac{\omega^4}{C^4}\right)}{4!}\cdots.
\end{aligned}
$$

This reduces for small values of l to

$$\cosh\delta = \cosh\left(l\sqrt{T^2 - \frac{\omega^2}{C^2}}\right).$$

Hence

$$\frac{p_n}{p_1} = e^{-nTl}e^{-n\delta} = e^{-nl\left(T + \sqrt{T^2 - \frac{\omega^2}{C^2}}\right)} = e^{-L\left(T + \sqrt{T^2 - \frac{\omega^2}{C^2}}\right)}$$

since $nl = L$, the total length of the horn.

As long as $T^2 > (\omega^2/C^2)$, an attenuation band exists, while if $\omega^2/C^2 > T^2$, the expression becomes

$$\frac{p_n}{p_1} = e^{-TL}\left[\cos\left(L\sqrt{\frac{\omega^2}{C^2} - T^2}\right) - i\sin\left(L\sqrt{\frac{\omega^2}{C^2} - T^2}\right)\right]$$

and a pass band occurs.

The complete equation for the horn, equivalent to equation (48), becomes

$$p_2 = e^{-LT}\left\{\left[\cosh\left(L\sqrt{T^2 - \frac{\omega^2}{C^2}}\right) + \frac{T}{\sqrt{T^2 - \frac{\omega^2}{C^2}}}\sinh\left(L\sqrt{T^2 - \frac{\omega^2}{C^2}}\right)\right]p_1 - \frac{iV_1\sqrt{P_0\gamma\rho}\,\frac{\omega}{C}\sinh\left(L\sqrt{T^2 - \frac{\omega^2}{C^2}}\right)}{S_1\sqrt{T^2 - \frac{\omega^2}{C^2}}}\right\},$$

$$V_2 = e^{+LT}\left\{\left[\cosh\left(L\sqrt{T^2 - \frac{\omega^2}{C^2}}\right) - \frac{T}{\sqrt{T^2 - \frac{\omega^2}{C^2}}}\sinh\left[\left(L\sqrt{T^2 - \frac{\omega^2}{C^2}}\right)\right]\right]V_1 - \frac{iS_1p_1\frac{\omega}{C}\sinh\left(L\sqrt{T^2 - \frac{\omega^2}{C^2}}\right)}{\sqrt{P_0\gamma\rho}\sqrt{T^2 - \frac{\omega^2}{C^2}}}\right\}. \quad (49)$$

These expressions can be derived from Webster's [9] differential equations for an exponential horn. Exponential horns have also been discussed by a number of writers.[10]

B. *Tapered Filters Whose Area Increases as the Square of the Distance*

One other example of a tapered filter, for which an approximate solution can be obtained, will be considered because of its bearing on the straight or conical horn. Let us assume that the area S_1 of a typical section of a tapered filter chain is n^2E, while that of the section next to it is equal to $(n + 1)^2E$, where E is a small constant. Sub-

[9] A. G. Webster, "Acoustic Impedance, and The Theory of Horns and of the Phonograph," *Nat. Acad. of Science*, Vol. 5, 1919, p. 275. The solution given by Webster for the exponential horn appears to have some typographical errors.

[10] Hanna and Slepian (*Trans. A. I. E. E.*, 43, 1924, p. 393); H. C. Harrison (British Patent No. 213,525, 1925); I. B. Crandall, "Theory of Vibrating Systems and Sound," D. Van Nostrand, 1926, p. 158.

stituting these values in equation (40), we obtain

$$\left[\left(\frac{(n^2+(n+1)^2)((n+1)^2+(n+2)^2)}{2(n)^2(n+2)^2}\right)\cosh\Gamma + \frac{n^2(n+2)^2-(n+1)^4}{2(n)^2(n+2)^2}\right] = K'\left(\frac{n^2+(n+1)^2}{2(n)^2}\right)e^{-\delta} + \left(\frac{(n+1)^2+(n+2)^2}{2(n+2)^2}\right)\frac{e^{\delta}}{K}.$$

If we substitute $K' = \sqrt{\frac{S_2}{S_3}} = \frac{n+1}{n+2}$ and $K = \sqrt{\frac{S_1}{S_2}} = \frac{n}{n+1}$ and neglect 1 as compared with n^3, we have

$$\cosh\delta = \left[\left(\frac{2n^2+1}{2n^2}\right)\cosh\Gamma - \frac{1}{2n^2}\right]. \tag{50}$$

If again our changes in areas are very small and hence n very large, we can neglect 1 compared with $2n^2$ and obtain

$$\cosh\delta = \cosh\Gamma, \quad \text{or} \quad \delta = \Gamma.$$

Either of these solutions will hold for any other pair of sections if we neglect 1 as compared with n^3 for the first of 1 compared with n^2 for the second. Hence for either solution, the propagation constant is little affected for this type of taper. The specific characteristic impedances Z_{R_1} and Z_{R_2} become

$$\begin{aligned} Z_{R_1} &= \frac{Z_0\sinh\Gamma}{\frac{1}{n}+\left(\frac{2n^2}{2n^2+1}\right)\sqrt{\left(\frac{2n^2+1}{2n^2}\cosh\Gamma-\frac{1}{2n^2}\right)^2-1}}, \\ Z_{R_2} &= \frac{Z_0\sinh\Gamma}{-\frac{1}{n}+\frac{2n^2}{2n^2+1}\sqrt{\left(\frac{2n^2+1}{2n^2}\cosh\Gamma-\frac{1}{2n^2}\right)^2-1}}. \end{aligned} \tag{51}$$

If we neglect 1 as compared with n^2, these expressions reduce to

$$Z_{R_1} = \frac{Z_0 n\sinh\Gamma}{1+n\sinh\Gamma}; \quad Z_{R_2} = \frac{Z_0 n\sinh\Gamma}{-1+n\sinh\Gamma}. \tag{52}$$

These impedances represent the impedances per square cm. looking in both directions at the input junction of the filter, whose area is n^2E. As we move in either direction these impedances change since n itself changes. If n becomes sufficiently large and Γ is not zero, the two characteristic impedances approach the value Z_0.

To express p_2 and V_2 in terms of p_1 and V_1 and these three parameters, we can obtain the equations.

$$p_2 = \frac{n}{n+1}\left[p_1 \left[\cosh\delta + \left(\frac{Z_{R_2}{}^I - Z_{R_1}{}^I}{Z_{R_2}{}^I + Z_{R_1}{}^I}\right)\sinh\delta \right] - \frac{V_1}{S_1}\left[\frac{2Z_{R_1}{}^I Z_{R_2}{}^I}{Z_{R_1}{}^I + Z_{R_2}{}^I}\right]\sinh\delta \right],$$

$$\begin{aligned} V_2 = \frac{n+1}{n}&\left[V_1 \left[\cosh\delta + \left(\frac{Z_{R_1}{}^O - Z_{R_2}{}^O}{Z_{R_1}{}^O + Z_{R_2}{}^O}\right)\sinh\delta \right] \right. \\ &- p_1 S_1 \left[\frac{Z_{R_1}{}^I + Z_{R_2}{}^I}{2Z_{R_1}{}^I Z_{R_2}{}^I}\right] \\ &\times \left[\sinh\delta \left[1 - \left(\frac{Z_{R_2}{}^I - Z_{R_1}{}^I}{Z_{R_2}{}^I + Z_{R_1}{}^I}\right)\left(\frac{Z_{R_2}{}^O - Z_{R_1}{}^O}{Z_{R_2}{}^O + Z_{R_1}{}^O}\right)\right] + \cosh\delta \right. \\ &\left.\left. \times \left[\left(\frac{Z_{R_2}{}^I - Z_{R_1}{}^I}{Z_{R_2}{}^I + Z_{R_1}{}^I}\right) - \left(\frac{Z_{R_2}{}^O - Z_{R_1}{}^O}{Z_{R_2}{}^O + Z_{R_1}{}^O}\right)\right]\right]\right], \end{aligned} \tag{53}$$

as can readily be seen by comparing these expressions with the equations, $p_2 = p_1 A - (V_1/S_1)Z_0 B$; $V_2 = V_1 C - (p_1/Z_0)S_1 D$. In the above expression the letter I indicates that the impedances Z_{R_1} and Z_{R_2} are to be taken at the input junction, while the letter O indicates that they are to be taken at the output junction.

The effect of this type of tapering is to change the propagation constant scarcely at all, but to lower the characteristic impedances in the neighborhood of the cut-off frequencies. This tends to produce large reflection losses and hence effectively the band is narrowed. A transforming action equivalent to a transformer of turns ratio $\sqrt{S_1/S_2}$ occurs as before.

To obtain the equation for a straight horn, we let S_1, a typical area of the horn, equal

$$S_1 = n^2 K = K'(nl)^2 = K'(x_1)^2,$$

where $nl = x_1$, the distance from the apex of the horn, and l the length of an individual section. Γ becomes $i\omega l/C$, and Z_{R_1} and Z_{R_2} are

$$Z_{R_1} = \frac{\sqrt{p_0\gamma\rho}\, in\frac{\omega}{C}l}{1 + in\frac{\omega}{C}l} = \frac{\sqrt{p_0\gamma\rho}\, i\frac{\omega}{C}x_1}{1 + i\frac{\omega}{C}x_1},$$

and (54)

$$Z_{R_2} = \frac{\sqrt{p_0\gamma\rho}\, i\frac{\omega}{C}x_1}{-1 + i\frac{\omega}{C}x_1}.$$

Substituting these values in equation (53), we obtain the equation

$$p_2 = \frac{x_1}{x_2}\left\{ p_1 \left[\cos\frac{\omega}{C}(x_2 - x_1) + \frac{\sin\left(\frac{\omega}{C}(x_2 - x_1)\right)}{\frac{\omega}{C}x_1} \right] - i\frac{V_1}{S_1}\sqrt{P_0\gamma\rho}\sin\frac{\omega}{C}(x_2 - x_1) \right\},$$

$$V_2 = \frac{x_2}{x_1}\left\{ V_1\left(\cos\frac{\omega}{C}(x_2 - x_1) - \frac{\sin\left(\frac{\omega}{C}(x_2 - x_1)\right)}{\frac{\omega}{C}x_2} \right) - i\frac{p_1 S_1}{\sqrt{P_0\gamma\rho}} \times\left[\left(1 + \frac{1}{\left(\frac{\omega}{C}\right)^2 x_1 x_2}\right)\sin\frac{\omega}{C}(x_2 - x_1) + \left(\frac{1}{\frac{\omega}{C}x_2} - \frac{1}{\frac{\omega}{C}x_1}\right)\cos\frac{\omega}{C}(x_2 - x_1)\right]\right\}. \quad (55)$$

If we introduce two lengths ϵ_1 and ϵ_2 defined by $\tan(\omega/C)\epsilon_1 = (\omega/C)x_1$ and $\tan(\omega/C)\epsilon_2 = (\omega/C)x_2$ and take account of the fact that the impedance as defined here must be multiplied by $i\omega$ to correspond to the impedance defined by Webster, then it is evident that the above equation corresponds to the relation given by Webster.[9, 10]

It is interesting to compare the relations obtained above involving the assumptions introduced in Section II with the solution involving no assumptions. This can be done for the conical horn, since its solution can be obtained using spherical waves. In Section III–*D*, the impedance looking into a conical horn was obtained when an infinite impedance terminated the horn. If we set $V_2 = 0$ in the last of equations (55) and solve for the ratio of p_1/V_1, it is evident that the impedance agrees with that given in Section III–*D*. Hence it is evident that both methods give the same solution.

Many other types of tapered filters can be solved in a similar manner, but no more will be considered here.

V. General Network Equations and Network Parameters

We can combine a number of symmetrical structures to form a general network. For any symmetrical structure we can write the

[10] The solution for the conical horn has been discussed in more detail by I. B. Crandall, "Theory of Vibrating Systems and Sound," D. Van Nostrand, 1926, p. 152.

equations

$$p_2 = p_1 \cosh \Gamma_1 - V_1 \frac{Z_{0_1}}{S_1} \sinh \Gamma_1,$$

$$V_2 = V_1 \cosh \Gamma_1 - \frac{p_1 S_1}{Z_{0_1}} \sinh \Gamma_1.$$

Suppose then that we wish to join this structure to other structures, with different characteristics and with different area conducting tubes. At the junction of the structures, we have by equation (14)

$$p_3 = p_2, \qquad V_3 = V_2.$$

Combining these with the above, we have

$$p_3 = p_1 \cosh \Gamma_1 - V_1 \frac{Z_{0_1}}{S_1} \sinh \Gamma_1,$$

$$V_3 = V_1 \cosh \Gamma_1 - \frac{p_1 S_1}{Z_{0_1}} \sinh \Gamma_1.$$

Writing a set of equations similar to the above for the second structure and combining, we have

$$\left.\begin{aligned} p_4 = p_1 &\left(\cosh \Gamma_1 \cosh \Gamma_2 + \frac{S_1}{S_2}\frac{Z_{0_2}}{Z_{0_1}} \sinh \Gamma_1 \sinh \Gamma_2 \right) \\ &- V_1 \frac{Z_{0_1}}{S_1}\left(\sinh \Gamma_1 \cosh \Gamma_2 + \frac{Z_{0_2}}{Z_{0_1}}\frac{S_1}{S_2} \cosh \Gamma_1 \cosh \Gamma_2 \right), \\ V_4 = V_1 &\left(\cosh \Gamma_1 \cosh \Gamma_2 + \frac{Z_{0_1}}{Z_{0_2}}\frac{S_2}{S_1} \sinh \Gamma_1 \sinh \Gamma_2 \right) \\ &- \frac{p_1 S_1}{Z_{0_1}}\left(\sinh \Gamma_1 \cosh \Gamma_2 + \frac{S_2}{S_1}\frac{Z_{0_1}}{Z_{0_2}} \cosh \Gamma_1 \sinh \Gamma_2 \right). \end{aligned}\right\}$$

We can also write this in the form

$$p_4 = p_1 \begin{vmatrix} \cosh \Gamma_1 & \frac{S_1}{S_2}\frac{Z_{0_2}}{Z_{0_1}} \sinh \Gamma_2 \\ -\sinh \Gamma_1 & \cosh \Gamma_2 \end{vmatrix} - V_1 \frac{Z_{0_1}}{S_1} \begin{vmatrix} \sinh \Gamma_1 & \frac{S_1}{S_2}\frac{Z_{0_2}}{Z_{0_1}} \sinh \Gamma_2 \\ -\cosh \Gamma_1 & \cosh \Gamma_2 \end{vmatrix}.$$

$$V_4 = V_1 \begin{vmatrix} \cosh \Gamma_1 & \frac{Z_{0_1}}{Z_{0_2}}\frac{S_2}{S_1} \cosh \Gamma_2 \\ -\sinh \Gamma_1 & \cosh \Gamma_2 \end{vmatrix} - \frac{p_1 S_1}{Z_{0_1}} \begin{vmatrix} \sinh \Gamma_1 & \frac{Z_{0_1}}{Z_{0_2}}\frac{S_2}{S_1} \sinh \Gamma_2 \\ -\cosh \Gamma_1 & \cosh \Gamma_2 \end{vmatrix}.$$

In fact if we combine η structures of this kind, we can write the equations

$$\begin{aligned} p_\eta &= p_1 A - V_1 \frac{Z_{0_1}}{S_1} B, \\ V_\eta &= V_1 C - \frac{p_1 S_1}{Z_{0_1}} D, \end{aligned} \tag{56}$$

where

$$A = \begin{vmatrix} \cosh \Gamma_1 & \frac{S_1}{S_2}\frac{Z_{0_2}}{Z_{0_1}} \sinh \Gamma_2 & -\frac{S_1}{S_3}\frac{Z_{0_3}}{Z_{0_1}} \sinh \Gamma_3 \cdots \\ -\sinh \Gamma_1 & \cosh \Gamma_2 & \frac{S_2}{S_3}\frac{Z_{0_3}}{Z_{0_2}} \sinh \Gamma_3 \cdots \\ \sinh \Gamma_1 & -\sinh \Gamma_2 & \cosh \Gamma_3 \quad \cdot \quad \cdot \\ \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \\ \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \\ \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cosh \Gamma_\eta \end{vmatrix} \tag{57}$$

$$B = \begin{vmatrix} \sinh \Gamma_1 & \frac{S_1}{S_2}\frac{Z_{0_2}}{Z_{0_1}} \sinh \Gamma_2 & -\frac{S_1}{S_3}\frac{Z_{0_3}}{Z_{0_1}} \sinh \Gamma_3 \cdots \\ -\cosh \Gamma_1 & \cosh \Gamma_2 & \frac{S_2}{S_3}\frac{Z_{0_3}}{Z_{0_2}} \sinh \Gamma_3 \cdots \\ \cosh \Gamma_1 & -\sinh \Gamma_2 & \cosh \Gamma_3 \quad \cdot \quad \cdot \\ -\cosh \Gamma_1 & \sinh \Gamma_2 & \sinh \Gamma_3 \\ \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \\ \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cosh \Gamma_\eta \end{vmatrix} \tag{58}$$

$$C = \begin{vmatrix} \cosh \Gamma_1 & \frac{S_2}{S_1}\frac{Z_{0_1}}{Z_{0_2}} \sinh \Gamma_2 & -\frac{S_3}{S_1}\frac{Z_{0_1}}{Z_{0_3}} \sinh \Gamma_3 \cdots \\ -\sinh \Gamma_1 & \cosh \Gamma_2 & \frac{S_3}{S_2}\frac{Z_{0_2}}{Z_{0_3}} \sinh \Gamma_3 \cdots \\ +\sinh \Gamma_1 & -\sinh \Gamma_2 & \cosh \Gamma_3 \quad \cdot \quad \cdot \\ \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \\ \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cosh \Gamma_\eta \end{vmatrix} \tag{59}$$

$$D = \begin{vmatrix} \sinh \Gamma_1 & \frac{S_2}{S_1}\frac{Z_{0_1}}{Z_{0_2}} \sinh \Gamma_2 & -\frac{S_3}{S_1}\frac{Z_{0_1}}{Z_{0_3}} \sinh \Gamma_3 \cdots \\ -\cosh \Gamma_1 & \cosh \Gamma_2 & \frac{S_3}{S_2}\frac{Z_{0_2}}{Z_{0_3}} \sinh \Gamma_3 \cdots \\ \cosh \Gamma_1 & -\sinh \Gamma_2 & \cosh \Gamma_3 \quad \cdot \quad \cdot \\ -\cosh \Gamma_1 & \sinh \Gamma_2 & -\sinh \Gamma_3 \\ \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \\ \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \; \cdot \; \cosh \Gamma_\eta \end{vmatrix} \tag{60}$$

Among these four determinants there is one relation

$$AC - BD = 1, \tag{61}$$

as can be seen by multiplying them together.

Hence to completely specify the characteristics of the structure three parameters are required. A number of possible sets of parameters exist whose usefulness depends on the type of structure to which they are applied. The set of parameters having the greatest use in connection with electrical networks are the image parameters which include two image impedances and an image transfer constant. We define these constants as follows for the acoustic case.

If we have a network terminated in impedances Z_{I_1} and Z_{I_2} (per square centimeter of area) at the beginning and at the end of the network, then these impedances are the image impedances of the structure if they terminate the structure in such a way that at either termination junction, the impedance looking in either direction is the same.

The image transfer constant θ may be defined as one half the natural logarithm of the vector ratio of the product of the pressure by the volume velocity, at the input junction point, and this product for the output junction point, when the network is terminated in its image impedances.

Hence

$$\theta = \frac{1}{2} \log_e \frac{p_1 V_1}{p_\eta V_\eta} .$$

To determine the image impedances, we have one set of equations

$$\left. \begin{aligned} p_\eta &= p_1 A - V_1 \frac{Z_{0_1}}{S_1} B, \\ V_\eta &= V_1 C - \frac{p_1 S_1}{Z_{0_1}} D. \end{aligned} \right\} \tag{62}$$

This gives the pressure and volume velocity propagated in one direction. We need also the equation of propagation in the opposite direction. This can evidently be written

$$\left. \begin{aligned} p_\eta' &= p_1' A' - V_1' \frac{Z_{0_\eta}}{S_\eta} B', \\ V_\eta' &= V_1' C' - \frac{p_1' S_\eta}{Z_{0_\eta}} D', \end{aligned} \right\} \tag{63}$$

where p_η' and V_η' represent the pressure and volume velocity at the beginning and V_1' and p_1' at the end of the structure. A' can be obtained from A by cyclically permuting the subscripts. By writing the expansions for these quantities we can show that

$$A' = C; \;\; C' = A; \;\; B' = \frac{Z_{0_1}}{Z_{0_\eta}} \frac{S_\eta}{S_1} B; \;\; D' \doteq \frac{Z_{0_\eta}}{Z_{0_1}} \frac{S_1}{S_\eta} D. \tag{64}$$

Eliminating the ratio V_η / V_1 from (62) and writing $p_1 / V_1 = Z_{I_1} / S_1$ and $p_\eta / V_\eta = Z_{I_2} / S_\eta$, we obtain

$$Z_{I_1} Z_{I_2} D + Z_{0_1} \left[Z_{I_1} \left(\frac{S_\eta}{S_1} A \right) - Z_{I_2} C \right] - Z_{0_1}^2 \left(\frac{S_\eta}{S_1} B \right) = 0. \tag{65}$$

From (63) eliminating the ratio V_η'/V_1' and writing $p_\eta'/V_\eta' = Z_{I_1}/S_1$ and $p_1'/V_1' = Z_{I_2}/S_\eta$ and substituting the values in (64), we have

$$Z_{I_1}Z_{I_2}D + Z_{0_1}\left(Z_{I_2}C - Z_{I_1}\frac{S_\eta}{S_1}A\right) - Z_{0_1}^{\ 2}\frac{S_\eta}{S_1}B = 0. \qquad (66)$$

Solving (65) and (66) simultaneously, we find

$$Z_{I_1} = Z_{0_1}\sqrt{\frac{BC}{AD}};\ \ Z_{I_2} = Z_{0_1}\frac{S_\eta}{S_1}\sqrt{\frac{AB}{CD}}. \qquad (67)$$

From the definition of θ and equations (62), we can show that

$$\cosh\theta = \sqrt{AC}. \qquad (68)$$

In terms of these parameters, the effect upon the pressure or volume velocity in the termination of an acoustic system, due to inserting the structure into the system, will be given by multiplying the terminal pressure or volume velocity by the factor

$$\frac{\dfrac{Z_A}{S_1} + \dfrac{Z_B}{S_\eta}}{\dfrac{2Z_B}{S_\eta}} \times \frac{2\sqrt{\dfrac{S_1}{S_n}}\sqrt{Z_{I_1}Z_{I_2}}}{Z_{I_1} + Z_A} \times \frac{2Z_B}{Z_{I_2} + Z_B} \times e^{-\theta}$$
$$\times \frac{1}{1 - \dfrac{Z_{I_2} - Z_B}{Z_{I_2} + Z_B} \times \dfrac{Z_{I_1} - Z_A}{Z_{I_1} + Z_A} \times e^{-2\theta}}, \qquad (69)$$

where Z_A and Z_B are respectively the impedances, per square centimeter, of the acoustic system at the insertion junction looking towards and away from the source.

APPENDIX I. PROOF OF THÉVENIN'S THEOREM FOR AN ACOUSTIC SYSTEM

The proof of Thévenin's theorem as stated in Section III can be obtained directly from the general network equations given in Section V. These equations are

$$p_2 = p_1 A - V_1\frac{Z_0}{S_1}B,$$
$$V_2 = V_1 C - \frac{p_1 S_1}{Z_0}D,$$

where $AC - BD = 1$. If we connect at the input end a source of pressure p_0, whose specific internal impedance is Z_T, we can write

$$p_0 = p_1 + V_1 \frac{Z_T}{S_1}.$$

Inserting this result in the above equation, we obtain

$$\left.\begin{aligned} p_2 &= p_0 A - \frac{V_1}{S_1}(Z_T A + Z_0 B), \\ V_2 &= V_1\left(C + \frac{Z_T}{Z_0}D\right) - \frac{p_0 S_1}{Z_0}D. \end{aligned}\right\} \tag{70}$$

To obtain the pressure when an infinite impedance is used at the termination, we let $V_2 = 0$, and solving for V_1 we have

$$V_1 = \frac{p_0 D S_1}{Z_0 C + Z_T D}. \tag{71}$$

Substituting this in the first of equations (70), we have

$$p_2 = \frac{p_0 Z_0}{(Z_0 C + Z_T D)} = p_0', \tag{72}$$

which is the terminal pressure for an infinite terminating impedance.

Eliminating V_1 from (70) and substituting $V_2 Z_R / S_\eta = p_2$, we have

$$V_2 = \frac{p_0 Z_0}{(Z_0 C + Z_T D)} \times \frac{1}{\dfrac{Z_R}{S_\eta} + \dfrac{Z_0}{S_1}\left(\dfrac{Z_T A + Z_0 B}{Z_0 C + Z_T D}\right)}. \tag{73}$$

We can show now that

$$\frac{Z_0}{S_1}\left(\frac{Z_T A + Z_0 B}{Z_0 C + Z_T D}\right) = \frac{Z_T'}{S_1},$$

which is the impedance at the terminating junction looking toward the source, when the specific impedance Z_T terminates the input end. From equations (63) and (64), we can write

$$p_2' = p_1' C - V_1' \frac{Z_{0_1}}{S_1} B,$$

$$V_2' = V_1' A - \frac{p_1' S_1}{Z_0} D.$$

Substituting $V_2'(Z_T/S_1) = p_2'$ and solving for the ratio p_1'/V_1', we have

$$\frac{p_1'}{V_1'} = \frac{Z_T'}{S_1} = \frac{Z_0}{S_1}\left(\frac{Z_T A + Z_0 B}{Z_0 C + Z_T D}\right). \tag{74}$$

Hence we can express V_2 in equation (73) as

$$V_2 = \frac{p_0'}{\frac{Z_R}{S_\eta} + \frac{Z_T}{S_1}}$$

which is Thévenin's theorem.

Appendix II. Determination of Loss for a Constant Volume Velocity Source

Another type of insertion effect desired in some cases is the effect caused by inserting filter structures in an acoustic system in which the source supplies a constant volume velocity. One such acoustic system is the phonograph.

In order to obtain this effect we first prove the theorem: If a source of constant volume velocity V_1 is connected to the input of an acoustic system, and if the impedance Z_R (per square centimeter) is used to terminate the system, the volume velocity V_2 will be $p_0''/[(Z_R/S_1) + (Z_c/S_\eta)]$ where p_0'' is the pressure at the termination of the system when the system is closed through an infinite impedance, and Z_c is the specific impedance of the acoustic system at the output junction looking toward the source when the system is terminated in in an infinite impedance at the input junction. S_1 and S_η are the areas at the input and output junctions, respectively.

To prove this we substitute the value of p_0 given by (71) in the first of equations (70) and obtain for the pressure, with an infinite impedance termination

$$p_2'' = \frac{V_1 Z_0}{S_1 D}. \tag{75}$$

Then eliminating p_0 from equations (70), and inserting the value $p_2 = V_2 Z_R/S_\eta$, we obtain

$$V_2 = \frac{V_1 Z_0}{S_1 D}\left[\frac{1}{\frac{Z_R}{S_\eta} + \frac{A Z_0}{S_1 D}}\right]. \tag{76}$$

From equation (74) we see that the impedance looking toward the source is $(Z_0 A/S_1 D)$ if we make Z_T approach infinity. Hence

$$V_2 = p_0''\left[\frac{1}{\frac{Z_R}{S_\eta} + \frac{Z_c}{S_1}}\right].$$

To obtain the insertion loss for a constant current source, then, it is only necessary to substitute Z_c for Z_a in equation (30). One special case of interest is the case where the acoustic filter is connected directly to the source. In this case $Z_c = \infty$ and the insertion effect is determined by the factor

$$\left(\frac{2Z_0}{Z_0 + Z_b}\right) \times e^{-\Gamma} \times \left(\frac{1}{1 + \dfrac{Z_0 - Z_b}{Z_0 + Z_b} e^{-2\Gamma}}\right). \qquad (77)$$

Contemporary Advances in Physics—XIII. Ferromagnetism

By KARL K. DARROW

Magnetism was revealed to Europeans by pieces of a mineral later to be called lodestone, which lay scattered in the fields of Magnesia in Asia Minor, and were endowed with the curious power of attracting iron. They who first noticed it were apparently Greeks of the period before the practice of writing; for legends of the discovery were transmitted by the Greeks of later centuries, legends entangled with tales of Cretan shepherds and the myth of Medea. Electricity was disclosed, evidently in the same dim period and region, by fragments of amber on which friction conferred the remarkable power of attracting shreds and flakes of light materials.

By these quaint phenomena electricity and magnetism were disclosed to the European world before the beginnings of written history; and the intimations were recorded in writings of classical antiquity, and handed down from generation to generation. Yet two millennia and more were destined to flow past, before sufficiently many further data should be gathered to make possible the forming of a valid conception of either. The nineteenth century arrived, before anyone detected the signs that the two are but different aspects of one fundamental entity. Obviously the early hints were not sufficient; but it would not be well to conclude that therefore the Greeks were unwise. If they are indicted for stupidity because they did not understand the lodestone and the electrified amber, the indictment lies also against ourselves. For these are instances of ferromagnetism and of frictional electricity; which is to say, they belong to provinces which to this day are not fully incorporated into the empire of the theory of electricity and magnetism.

How then does it happen that the phenomena earliest discovered must still be listed among the least well understood? There is nothing unusual in this. There is no general reason for expecting that the phenomena which occur spontaneously and frequently and conspicuously in Nature should be the easiest to understand. On the contrary, it frequently happens that they are much less instructive and interpretable than others which can be brought to pass only by careful choice of conditions and skilful experimentation. The history of physics abounds in instances of such contrasts, and there is none more striking than the one with which I am to deal. Many phenomena of magnetism are well explained by the contemporary theory, many seem admirably clear; but none of these was or could have been witnessed by the Greeks. We know much about the magnetic properties of gases, dilute solutions, free atoms, elements and compounds which are so feebly magnetizable that before 1830 they were not supposed to be "magnetic" at all; we are still perplexed by the behaviour of iron and lodestone. This is the reason why there are textbooks of magnetism, in which hundreds of pages are devoted to the data and the theories of a number of effects most difficult to perceive and known to none but physicists, while the magnets of daily experience are dismissed with a chapter or two of mere description. As for electrified amber and its kindred, they are fortunate to have a few paragraphs of any modern treatise on electricity bestowed upon them.

Frictional electricity is not a very striking phenomenon, nor is it valuable in engineering; consequently it has been allowed to slip into obscurity, shunned by cautious students on the hunt for problems promising immediate returns. Ferromagnetism is not so unobtrusive. Much of the electric machinery which has transformed the world since Napoleon derives all its efficacy from certain blocks of iron or magnetizable alloy, enmeshed among the wires. So useful a property of matter does not consent to lie neglected; physicists are forced to hearken to its insistent demands for attention. Ambition to achieve some technical advance supplies a strong incentive; and there is a feeling of humiliation that a quality of matter so conspicuous and so remarkable, and so remarkably limited to a particular class of substances not in other ways exceptional, should not be properly connected with the structure of contemporary physics. For these and other motives, there are always physicists engaged in the struggle with the problem of ferromagnetism—no mean struggle, for the difficulties are truly serious. It was a tough problem which was offered to the Greeks and which they rejected, when they saw the lodestone, took note of it, and left it for the modern world to study.

Some of the difficulties of ferromagnetism may be peculiar to it. Others, it is to be feared, are examples of the troubles which are reserved for scientists by the internal properties of solid bodies generally, and which physicists will some day be forced to confront when the obvious problems of gases and free atoms are exhausted, if they are not sooner incited by curiosity or by the requirements of engineering. Most of the great conquests of recent physics have been achieved through the study of gases, or of those properties of matter which are the same for the solid as for the gaseous state. It is but natural to wish to postpone as long as possible the attack upon the intrinsic properties of solids; but there is no evading it in the study of ferromagnetism, for this is a property of solids only, and not even of transparent solids at that. One would wish at least to be permitted to restrict the study to pure elements or simple compounds; but many of the most interesting of the ferromagnetics belong among those bewildering substances the alloys, which form what the mathematicians describe as a *continuum* beside and among the great yet finite number of chemical compounds. If one were to work only with perfectly pure iron (supposing that one could get such a substance, or could recognize it when he had it!) the problem would not yet be simple; for every species of mechanical and thermal treatment, and magnetization itself, would transform the iron into a new sort of solid.

These difficulties I will stress in the pages of this article. There is another. The information about a ferromagnetic substance—the prime material required for theorizing or for practical applications—is usually furnished in the form of so-called I-vs-H curves; that is to say, relations between the "intensity of magnetization" and the "magnetizing field." These curves play the part of the ultimate data of experience. Yet they are not ultimate data; the "magnetizing field" is seldom actually measured, the "intensity of magnetization" almost never. These entities I and H are deduced from experience by means of a theory. The theory is indispensable. If an uninstructed person were presented with a number of variously-shaped pieces of iron, and a battery and a coil of wire with which to produce any desired magnetic field, and any number of measuring-instruments, he would find it extremely hard to select something to measure that might yield a coherent and intelligible set of data. He would be able to show in a vague way that the greater the magnetic field acting upon any piece of iron, the more powerful a magnet it becomes; but if he were to search for some precise measurable quantity that could serve as a measure of the power of the magnet, and that would be *characteristic of iron as a substance and not merely characteristic of individual pieces of iron as individuals*, his search would be a long one. From what I have just called "the theory" he would find out what to measure, and how to calculate from it the value of something characteristic of iron and not affected by the shape of the pieces; he would find out how to trace an "I-vs-H" curve. This curve would serve in turn as a basis for theories of ferromagnetism; but theory would have entered already into the preparation of the curve. I shall therefore devote the first section of this article to the principles according to which such curves are determined from the immediate data. Any reader who feels that these principles are familiar, or self-evident, or unimportant, may leap to the second section and the third, in which the I-vs-H curves are accepted as the data of experience.

I shall not venture a definition of ferromagnetism until nearly the end of the article. Such a definition is not easy to make, unless one takes refuge in the statement that "ferromagnetism is the kind of magnetism displayed by iron." I can only regret the frequency with which such ponderous words as *ferromagnetism* and *permeability* and *susceptibility* and *magnetization* and *magnetostriction* must needs appear. The subject is encumbered by its heavy vocabulary; it ought to have a new one made up entirely of short and vivid words.

A. Analysis of the Magnetization of Magnetized Bodies.

Let us imagine a collection of magnets such as one frequently sees, horseshoe magnets for example, with their ends painted red and blue. We know that (if the painting was done properly) the red end of each attracts the blue ends and repels the red ends of the others; the blue end attracts those which the red end repels and repels those which the red end attracts. It seems as if the ends of the magnets were covered with invisible substances—one kind on all the red ends, the other on all the blue ends—so constituted, that a sample of either substance attracts all samples of the other sort, repels all samples like itself. Coulomb found that if the magnets were long and slender, so that the power of attracting and repelling was concentrated very closely about the extremities of each, these extremities attracted or repelled one another according to an inverse-square law. That suggested gravitation and electric force; which suggested in turn that, even as matter is the source of gravitation and electric charge is the source of electric force, so also there is an invisible thing called magnetism which inhabits iron—or rather, two invisible things, positive magnetism and negative magnetism, which may be pulled and pushed around inside and over the surface of a piece of iron. This notion of a pair of invisible and mobile fluids is very helpful, and I shall use it in several passages; yet the reader must not take it as corresponding to the actual reality. We cannot imagine two or even one perfectly mobile magnetic fluid, for a well-known reason.

The reason is, that even though a magnet may appear to carry nothing but positive magnetism on one of its ends and nothing but negative magnetism on the other, yet it is not possible to cut off anywhere a piece containing only one of these kinds. In fact it is not possible anywhere to cut off a piece not containing equal quantities of the two kinds of magnetism. Any piece of matter always contains as much positive magnetism as negative; so also does any smaller fragment broken off from the piece, and any still smaller bit broken out of the fragment, and so forth until the original piece is crumbled into dust, each particle of which still contains as much magnetism of either sign as of the other.*

Now this requires that when we subdivide a magnetized piece of iron into tiny parcels or volume-elements, not by the hammer nor the file but by the exercise of the imagination, these volume-elements must themselves be imagined as magnets each invested with a positive pole and a negative pole and a magnetic axis pointing in some particular direction. I am not implying atoms by these "parcels"—we shall as yet have nothing to do with atoms. The process of dividing a substance into imaginary small volume-elements has nothing in common with the construction of atoms or atom-models; quite the contrary! It is a process which every physicist undertakes, whenever he desires to analyze the flow of water or the vibrations of air or the strain of a twisted rod or any of a multitude of problems concerning pieces of matter, which, whatever his views about atoms, he intends to regard as continuous media for the nonce. Well! in dealing with magnetism, it is not sufficient to conceive these volume-elements as cubical or otherwise-shaped bits of matter entirely uniform and isotropic in their qualities; they must be conceived as being little magnets themselves.

This is the reason why we are taught to imagine a piece of magnetized iron as a collection of tiny cubes, each bearing positive magnetism spread like a coat of paint over one side, and negative magnetism over the side opposite; or as a bundle of filaments which, where they come out to the surface of the piece, divide it into a pattern of area-elements each of which is overspread with magnetism positive or negative; or as a pile of laminæ, somewhat like a nest of saucers, each of which is covered with magnetism of the two signs on its two sides. This is the reason why, developing the first of these conceptions (which contains implicitly the other two), we are taught to picture a function called the *intensity of magnetization*, which has a definite value at each point within the magnet, and may be visualized with the aid of the imaginary cubes. Select a point in the interior of the magnet, and imagine it surrounded by a cubical volume-element of thickness d and face-area d^2 and volume d^3; and imagine two opposite sides of the cube to be covered with magnetism of opposite signs painted on with a surface-density I, so that each side bears a quantity Q which

* The best evidence for this statement is the fact that magnets in a uniform magnetic field such as that of the earth experience no force tending to displace them bodily though they experience a torque tending to orient them.

is Id^2. This cube would be a minute magnet having the moment* Qd which is Id^3, directed normally to the two sides coated with magnetism; for I is a quantity possessing both magnitude and direction, a vector quantity and not a scalar—this is a way of expressing the complexity to which an allusion was made in the last paragraph. The piece of magnetized material is to be visualized as the assembly of all these little cubes, each having a magnetic moment equal to its volume multiplied into the value of I prevailing in it. The force exerted by the piece anywhere outside of its volume is to be considered as the sum of the forces there exerted by all the little magnets. The entity I plays the rôle of a magnetic-moment-per-unit-volume. It is this entity which is defined as the *intensity of magnetization* of the material.

This, it may be objected, is something quite unverifiable; for one cannot penetrate into the interior of a piece of iron, and find out whether it contains such an entity as this vector I. Quite so! and this is another of the great difficulties in ferromagnetism, though not peculiar to ferromagnetism alone, for it besets in greater or less degree every problem of the properties of solid bodies. The state of affairs within a piece of magnetized iron is the leading problem of ferromagnetism, indeed it is the one problem which contains all the rest. But there is no way of ascertaining that state of affairs, for there is no way of putting a measuring-instrument into a piece of iron. One might scoop a hole in the iron to make a place for the magnetometer, but then the magnetometer would be in the hole and not in the iron. The field of magnetic force outside the magnet can be plotted, the lines of force in the field can be followed up to the very edge of the magnetized material, but there they dive and they disappear. When one sees a sketch of a magnet and its environment, in which the lines of force coming up from all sides to the surface of the magnet are connected in pairs by "lines of induction" passing through the body of the magnet, he should realize that while the lines of force outside are a map of a field which can be explored, the lines of induction within are hypothetical altogether.

Why then take the trouble of conceiving entities such as these, intensity of magnetization I and induction B, since they are solely imagined to exist in a locality where there is no possible means of penetrating to seek them? The reason is this, and this only: Confined though they are within the bodies of the magnets, they facilitate the

* "Magnetic moment" is usually defined by inviting the reader to imagine a magnet so long and slender that its "magnetism" is concentrated almost completely at its ends or "poles"; the moment of such a magnet is the product of its length into the amount of magnetism, or "polestrength," at either end. Actual magnets have no true poles. The moment of an actual magnet is the torque which a unit field exerts upon it when it is normal to the direction in which the field tends to set it.

understanding of the effects which the magnets produce outside. Induction and intensity of magnetization are things which are supposed to exist *inside* a solid magnetic body, to make it possible to predict what effects that body produces in the world outside of itself—the only region which can be entered with or without measuring-instruments.

Now if a magnet were delivered over by Nature in fixed and permanent state, so that nothing which could be done to it would alter its behavior towards surrounding objects, the problem of determining I would be relatively simple. It would amount to this: to build up a structure of little cubical magnets occupying the same volume as the actual magnet, and producing everywhere outside that volume the same field as the actual magnet is observed to produce. In other words, it would consist in seeking a function I of the coordinates x, y, z of the points within the volume of the magnet, fulfilling the following condition: when this volume is subdivided into small cells of volume dv, and each is treated as a magnet of moment Idv, and the forces exerted by all these little magnets at any point outside of the volume are summed together, their sum shall turn out to be the same as the force which the actual magnet is observed to exert at that point.

This however is not the whole of the actual problem. The force which a magnet exerts at any particular point in its vicinity depends upon the magnetic fields which are impressed upon it by external objects—other magnets, or electric currents, or the earth itself. It becomes a different magnet when it is subjected to a different field. The process of finding a function I fulfilling the condition made above must therefore be carried through anew whenever the exterior fields acting upon the magnet are changed.

This variability makes the problem much more difficult. Yet in some cases it can be dealt with, in the same manner as the more restricted problem of analyzing an unchanging magnet into volume-elements; and in dealing with it, the first foundations of a theory of magnetism are laid down.

A piece of iron is observed to become a different magnet, whenever the impressed magnetic field is changed. Very well! we will try to describe the difference, by assuming that each of the volume-elements into which we have mentally divided the piece becomes itself a different magnet. The change in the magnetism of the piece is all too likely to be complicated and obscure; but we will simplify by supposing that the magnetization of each of the volume-elements depends upon the magnetic field prevailing in it, according to some law which is the same for all the volume-elements; that there is a fixed relation between the intensity of magnetization at a point and the field existing at that point, which is the same everywhere within the supposedly uniform piece of iron, which is a quality of that particular kind of iron. If there is no such relation, the whole procedure is likely to be futile. If there is such a relation, it is the fundamental fact of magnetism; and the first business of the student of magnetism is to determine it for as many substances, under as many conditions, as he can. We shall presently see that most research in ferromagnetism is devoted to determining this relation, by methods which would not yield self-consistent results, did it not exist.

But we shall attain nothing by merely assuming that there is such a relation, unless we make another assumption concerning the field prevailing within the magnet; for it is quite inaccessible, we cannot enter in to measure it. Let us therefore suppose that the field produced at any point inside the magnet, by the objects outside—be they laboratory magnets, or electric currents, or the earth itself—is the same as they would produce at that point were the magnet taken away, leaving them the same. The outer parts of the magnet are supposed *not* to shield the inner parts from the magnetic influences of the outer world. This is a natural corollary of the supposition we have tacitly made already, that the outer volume-elements of the magnet do not shield the outer world from the magnetic forces due to the inner volume-elements. We assume it; and we assume that the intensity of magnetization and the magnetic field, the vectors I and H, are parallel to one another,* and that there is a relation between their magnitudes which is the same for every point within the magnet.

On proceeding to test this set of assumptions by the appeal to experiment, we encounter results which at first sight seem to destroy them. For instance, let us immerse a short rod of iron (quite demagnetized to begin with) in the uniform magnetic field produced within a long cylindrical tube by an electric current flowing through a coil of wire, a solenoid, evenly wrapped around the tube. The field H_e which the current would produce within the tube were the iron not there is uniform in magnitude and direction, everywhere parallel to the axis of the solenoid. By the last assumption, this is the field which the current produces everywhere inside the iron. We map the magnetic field produced by the rod in its vicinity, and determine

* There are cases, neither few nor trivial, in which I and H cannot always be supposed parallel; for instance, when the magnet is a large crystal, or when it is a plate of metal which has been cold-rolled, or when the direction of the magnetizing field is changed after the substance is already perceptibly magnetized. But if I were to expound the most general actual case, this article would never come to an end.

the function I which describes the magnetization which would produce such a field. The vector I is not uniform throughout the iron, either in direction or in magnitude. Though H_e is the same everywhere within the metal, I varies from point to point. This result by itself seems to demolish the assumptions.

The contradiction however is only apparent; it vanishes if we make due allowance for the field produced at every part of the magnet by the other parts, for *the effect of the magnet upon itself*. Continuing to use the illustration of the short rod in the uniform impressed field: the distribution of elementary magnets which the function I expresses, and which produces at every point outside the iron a calculable field agreeing with the field there observed, should also produce a calculable field at every point within the iron. Considering that we have assumed that the force due to even the innermost volume-element of the magnet is exerted unimpeded everywhere in the outside world, we cannot consistently avoid assuming that its force is exerted unimpeded upon the other volume-elements as well. Thus it is reasonable to suppose that if the value of I at any point in the iron is controlled by the magnetic field there prevailing, then the truly controlling field comprises not only the one (H_e) due to the external agencies, but also the other (H_i) due to the multitude of little magnets presumed to constitute the piece of iron. The value of I should depend on the resultant H of H_e and H_i. In the present case of the short rod inside the solenoid, the vector H_e is uniform, but the vector H_i varies from point to point, and consequently so does the resultant H of H_e and H_i, and consequently so does I. More properly, I should not use such a word as "consequently" at all; both I and H vary from point to point, either accounting for the other, either being cause and either being effect.

This, by the way, is one of the reasons why as a rule it is not possible to analyze the magnetization of a magnet by cutting it into little pieces and measuring the moment of each separately. When such a piece is isolated from the rest of the magnet, the field acting upon it is changed even though all the external field-producing agencies remain the same. The other reason for not cutting up a magnet is, that the stresses exerted on the material in the process of cutting are likely to change it into some very different ferromagnetic material—but of this, more later.

The problem of determining I now assumes its full scope. For every magnet, or let us say for every piece of magnetized iron, there should be a function I describing its magnetization, defined at every point within it and satisfying these conditions:

First, it should account for the field due to the magnet at every point outside;

Second, its value at every point inside the magnet should be a definite function of the thing which we have just tentatively defined as "the magnetic field" at that point; *viz.* the resultant of that field which the external agencies would produce were the magnet away, and that which the magnetization I should itself produce.

Or, in other words: it should be possible to build up a reproduction of the magnetized piece of iron out of little magnets, the magnetic moment of each depending in a perfectly definite way on the force exerted on it by the other little magnets and by the external world, and all together producing the same effects in the external world as the piece of iron does.

In saying "it should be possible" I do not mean to imply that there is an obligation resting upon Nature to construct magnetizable objects in such a way that it is possible. One could not prove *a priori* that she does. One must take variously shaped pieces of magnetizable metal and observe their behavior in various impressed fields, and ascertain for each whether or not there is a function I. In so doing, one is liable to encounter very great mathematical difficulties. In fact, the difficulties are likely to prove insuperable unless the piece of metal is shaped in one or other of a few definite ways, and the impressed field is uniform and properly oriented.

Let us attack the problem from the other side, and enquire first whether it is possible so to shape a piece of iron and so to orient the impressed field, that the extra field due to the magnetization should vanish everywhere within the iron, and the actual field should everywhere be identical with the impressed field—so that although there is a function I differing from zero, yet $H_i = 0$ and $H = H_e$ everywhere inside the iron. This condition would be realized, if one could make an infinitely long straight rod and expose it to an infinitely extended uniform field parallel to its axis. It is very nearly realized along the middle of a wire several hundred times as long as it is thick, set parallel to the earth's field or along the axis of a solenoid somewhat longer than the wire itself. It is very nearly realized within the substance of a ring-shaped piece of metal pervaded everywhere by an impressed field following the curvature of the ring; a field of this character can be produced by wrapping a current-carrying wire around the ring.

In these cases, or rather in the ideal cases to which these are close approximations, the vectors H_e and I are uniform throughout the metal; the relation between their magnitudes is the relation between "magnetizing field" and "intensity of magnetization," which is characteristic of the metal and is the cardinal fact of ferromagnetism.

Next we enquire whether it is possible so to shape the metal and so to orient the impressed field, that the actual field within the metal shall be uniform all through it even though not the same as the impressed field—so that I and H_e and H_i shall all three differ from zero, and the resultant H of H_e and H_i shall be uniform throughout the magnet. This condition is realized, if the piece of metal is an ellipsoid and the impressed field is uniform and directed parallel to one of its axes. In this case the ellipsoid is magnetized uniformly, and the extra field H_i which it produces within itself is uniform and oppositely directed,* "antiparallel," to the impressed field. The actual field H is uniform and points everywhere in the same direction as H_e, and its magnitude is equal to the difference between the magnitudes of H_e and H_i. The magnitude of H_i is proportional to that of I, as might be expected, so that

$$H = H_e - NI.$$

The factor N ("demagnetizing factor") depends upon the ratios between the axes of the ellipsoid, and Maxwell developed formulæ for it.

In these cases of ellipsoids, the relation between I and H_e, which is what the data usually supply, is not the true relation between the intensity of magnetization and the magnetizing field. However, the more significant relation between I and $H_e - NI$ can be deduced from the other by a simple graphical artifice. Ellipsoids of different shapes yield very different I-vs.-H_e curves; but the I-vs.-H curves into which these are translated in the aforesaid manner agree with one another, and with the curves obtained from closed rings or exceedingly long wires, very well indeed. Did they not agree, the whole theory would be upset; this procedure therefore is a manner of testing the theory.†

Incidentally, the field H_i produced by the magnet within itself may be far from insignificant. To take an example from Ewing:

* Unless the metal was not properly demagnetized before the application of the field.

† The artifice mentioned above consists in drawing upon the graph, on which orthogonal axes for I and for H_e have already been laid off, an additional axis passing through the origin and inclined to the I-axis at an angle of which the tangent is N. If now the I-vs.-H_e curve is plotted in the usual way, the value of H corresponding to any point P upon the curve is given by the length of the line drawn parallel to the H_e-axis and connecting P with the new axis.

In dealing with rods or other magnets shaped differently from ellipsoids, N may be determined empirically by plotting the I-vs.-H_e curve and drawing an axis so inclined to the I-axis that when the curve is referred to the new axis it coincides with the curve obtained with an ellipsoidal or ring-shaped magnet of the same material; the value of N is then the tangent of the angle between the new axis and the I-axis.

inside a sphere of soft iron exposed to the earth's magnetic field, H_i amounts to 84/85 of H_e, so that only 1/85 of the external field is active within the iron. Since the discovery of permalloy, this instance can be bettered. Within a sphere of suitably prepared permalloy exposed to a field of 10,000 gauss, 0.9996 of that field is counteracted by the magnetized volume-elements themselves.

This counterbalancing of part of the impressed field is sometimes called the *demagnetizing effect of the poles*—a rather unfortunate term, which affords me a pretext for discussing these alleged "poles." The pole of a magnet is like the end of the rainbow; if one were to tunnel into a magnet to get the pole, one would not find it. Or, to draw a better simile from geometrical optics, the poles of a magnet are like virtual images behind a mirror. The virtual image is a point which we reach by retracing the light-rays backward to the surface of the mirror and then prolonging them straight ahead until they all intersect, even though the light-rays themselves actually came up to the mirror from some other direction; the magnet-pole is a point which we reach by prolonging the lines of force down into the substance of the magnet and carrying them on until they meet, although the lines of force actually supposed to prevail within the magnet may not converge at all. The poles, in fact, are like all the other entities supposed to exist inside a magnet—they are imagined, in order to describe and predict the field which the magnet produces outside of itself. For instance, the external field due to an ellipsoid magnetized parallel to an axis is precisely that which two "poles," properly placed upon the axis and endowed with the proper equal amounts of positive and negative magnetism, would produce. If one chooses to visualize these "poles" rather than the ellipsoid, there is nothing to impede him.*

Again it is permissible, in the case of the ellipsoid and in some others, to visualize only the "magnetization of the surface"—to imagine the surface painted over with magnetism, laid on with a density governed by a certain law. At any point P on the surface of the ellipsoid, let $\mathbf{I}$ represent the magnetization of the material, which as we have seen is a vector; let I stand for the magnitude of this vector; let ds stand for the area of a small element of the surface containing P; let θ stand for the angle between the outward-pointing

* The inexactitude of this concept of "poles" leads to some curious lapses of logic in most expositions of the theory of magnetism (including, I am afraid, this one). Even in Maxwell we read: "The ends of a long thin magnet are commonly called its poles. . . . In all actual magnets the magnetization deviates from uniformity, so that no single points can be taken as the poles. Coulomb, however, by using long thin rods magnetized with care, succeeded in establishing the law of force between two like magnetic poles."(!)

Some use the terms "poles" or "polestrength" in the sense assigned to the word "magnetism" on p. 298.

normal to ds and the vector **I**. Magnetism in the amount $I \cdot ds \cdot \cos \theta$ is to be spread upon ds; magnetism is to be spread over the surface of the ellipsoid with surface-density $I \cdot \cos \theta$. This film of "magnetism" would produce, everywhere outside of the ellipsoid, the same field as the poles or the continuous magnetization which we have imagined as existing inside the ellipsoid. Furthermore, it has a firmer basis in experience than do the poles. For, if a beam of polarized light is directed against the surface of a magnetized ellipsoid, the reflected beam is curiously altered; this effect, known by the name of its discoverer Kerr, is sometimes extremely complicated, but in magnitude it is always proportional to the value of the imagined quantity $I \cdot \cos \theta$ at the point where the reflection occurs; and by promenading a spot of light over a magnetized piece of iron and analyzing at every point the reflected beam, one can actually find how $I \cdot \cos \theta$ varies all over the surface. This property endows the vector I with a physical reality.

There is still one of the effects which a magnet produces outside of itself, which requires our attention; did it not exist, magnets would not play nearly so great a rôle as they do in the life of the world.

Hitherto I have implied that one maps out the external field of a magnet by exploring it with some one of the known field-measuring devices, of which there are several: the magnetometer needle, the bismuth wire which changes its resistance according to the field impressed upon it, the plate of glass which rotates a traversing beam of plane-polarized light to an extent proportional to the field. There is another method essentially different from these, and capable of measuring something which they cannot. One may set up a loop of wire in the neighborhood of the to-be-magnetized piece of metal; suddenly impress the magnetizing field; and measure the sudden rush of charge around the loop. This rush of charge is proportional to the mean value of the magnetic field thus suddenly created in the region enclosed by the loop.* One might map a field in this manner; but that is not the unique feature of the method.

We consider a special and actual case. Take an unmagnetized ring of iron; cut out a thin segment, leaving two nearly parallel end-surfaces facing one another across a narrow gap; take a loop slightly larger than the cross-section of the ring, suspend it in the gap,

* The E.M.F. around the loop at any instant is equal to the time-derivative of the surface-integral, over any surface bounded by the loop, of the component of the magnetic field normal to the surface; in technical language it is equal to the rate of change of the flux of magnetic field through the loop. The rush of charge is equal to the quotient of the time-integral of this E.M.F., which is the difference between the initial and final values of the surface-integral, by the resistance of the loop.

parallel to the end-surfaces; apply an impressed field H_e by sending a current through a coil wrapped around the ring. The rush of charge in the loop testifies that the field established in the gap is vastly greater than H_e, a fact which can be confirmed by the magnetometer or any other field-measuring device. The field in the gap is, in fact, the resultant of H_e and a field due to the magnetized iron. We call it B. Replace the segment, closing the ring; encircle the restored segment with the loop as with a collar; repeat the experiment (after carefully demagnetizing the ring, so as to start afresh from the same condition). The rush of charge is the same. The apparent inference is, that the field B continues to subsist inside the iron forming the closed ring; and the method of the loop seems to be competent to measure, if not the actual force within the metal, at least the average of its values—which would contradict in part my former statement that the field within the iron is unreachable by measurement.

The contradiction involves one of the most confusing assumptions in the theory of ferromagnetism. The field B is greater than the impressed field H_e, whereas the actual field H, which we have been postulating within the iron in order to explain its magnetization, is smaller than H_e. To prove this for the ring might be difficult, since it is a property of the complete ring that the field due to its own magnetization is zero everywhere outside of it as well as inside (so that, incidentally, the method of the loop is the only one giving even an intimation that the ring is a magnet). With an ellipsoid the demonstration is easy. Wrap the loop like a girdle around the middle of an ellipsoid of iron, and suddenly magnetize the iron by impressing a uniform field H_e parallel to one of its axes and normal to the plane of the loop. Measure the rush of charge; it attests that the field established through the loop is much greater than H_e. But the field within the iron, as we have seen already, has been set equal to $H_e - NI$, hence to a value smaller than H_e, in order to account for the field outside.

It is necessary, therefore, to add a third vector B to the pair of vectors I and H which we have already conceived as existing in the depths of the magnet. It is this vector, the alteration of which governs the rush of charge which occurs through a loop encircling the magnet when the magnetization is changed. The rush of charge is proportional to the change in the mean value of B throughout the magnet in the plane of the loop—not to the mean value of H. Making this the definition of B, and considering all the data assembled from experiments on rings and ellipsoids and rods of various proportions, it is found that the observations made upon their external fields by field-measuring devices and the observations made by the method of the loop are all reconcilable with one another, provided that the vector B is made parallel to I and H and equal to

$$B = H + 4\pi I.$$

The vector B is known as the *induction*. The relation between B and H is often plotted instead of the relation between I and H; naturally if either relation is known the other can readily be found. The ratio of the magnitudes of B and H is called *permeability* and denoted by μ; the ratio of the magnitudes of I and H is called *susceptibility* and denoted by κ or σ or χ.

One might think that this quantity B should be identified with the magnetic field which is supposed to exist within the metal and to magnetize it. Though all the textbooks beseech the student not to confuse the induction with the field (he is usually asked to imagine himself digging variously shaped infinitely small holes within a magnet, and putting an instrument into each to measure the magnetic force inside it), the distinction has an obstinate way of not becoming clear. We should get just as self-consistent sets of curves if we were to plot I against $(H_e + H_i + 4\pi I)$ as we do when plotting I against $(H_e + H_i)$; it would merely be tantamount to adding 4π to the "demagnetizing factor." As a matter of fact nearly everyone, as soon as he begins to theorize about the state of affairs inside magnetized bodies (or polarized dielectrics), promptly assumes that the acting field is something different from the resultant of H_e and H_i. Some make it equal to $(H + \frac{4}{3}\pi I)$, attributing the term $\frac{4}{3}\pi I$ to an action of the molecules which are neither very close to nor very far from the point where the field is being evaluated. Some (Weiss and his many followers) make it equal in ferromagnetic metals to the sum of H and a term nI, the factor n being so enormous that the postulated field is millions of times as great as H and thousands of times as great as B. The extra field, they say, is "not magnetic"; but this distinction is more obscure than the other. Nobody really knows what the field inside a magnetized solid is. The best policy is to continue plotting I and B as functions of H, regarding H as the independent variable sanctioned by tradition.

B. The Relation between Intensity of Magnetization and Magnetic Field

Since all of the actions of magnets are interpreted by supposing that in every magnetizable substance the intensity of magnetization is controlled by the magnetic field in a definite and peculiar way—

that for every magnetizable substance there is a distinctive I-vs.-H relation—it is evident that this relation, if it exists, must be the fundamental fact of magnetism. The first object of research in ferromagnetism is to discover it for all of the ferromagnetic materials; the second, to devise for each of these materials a model, accounting for the particular form of I-vs.-H relation which it displays.

On setting about to collate the recorded samples of I-vs.-H curves, one promptly encounters the last and greatest of the troubles of ferromagnetism. There are infinitely many such curves to be collected, for there is a limitless variety of ferromagnetic substances!

This is not always realized, because of the unfortunate practice of referring to "the three ferromagnetic metals, iron, nickel and cobalt," as though there were but three I-vs.-H relations to be determined. But in addition, there are ferromagnetic alloys: binary alloys of iron with nickel, of nickel with cobalt, of cobalt with iron; ternary and yet more complex alloys comprising these and other elements, or consisting entirely of elements none of which by itself is ferromagnetic. Anyone acquainted with the diversities of alloys would be prepared to find a truly vast variety of qualities exhibited by these; and he would not be disappointed. Indeed, an alloy may contain one of its elements in so small a proportion as to appear quite negligible—so small, as to be considered a mere casual impurity—so slight, as to be difficult to detect and difficult to expel—and yet so great, as to influence the magnetization in the most drastic fashion. Iron containing a fraction of a per cent of carbon differs as much from pure iron, in regard to its magnetic properties, as either differs from nickel. (Perhaps even what is now called "pure" iron contains a minimal amount of some undetected yet potent impurity, the ultimate removal of which will reveal a whole new set of phenomena!) So there is not a triad, but a multitude of ferromagnetic substances, each of which may be expected to have a distinctive I-vs.-H relation of its own.

But for each of these substances there is, as it turns out, not one but a legion of I-vs.-H relations. The curve depends very much on the temperature of the sample—to such an extent, indeed, that as the temperature is raised, the ferromagnetism varies rapidly, diminishes, and finally vanishes. The curve depends also upon the mechanical stresses prevailing in the material, compression and tension and torsion and the complicated combinations of these. It is also liable to be altered by an electric current flowing in the material.

Degree of crystallization likewise matters a great deal. Most of the samples of metal used in the past have consisted of very great numbers of very small crystals, millions of them to a cubic centimeter.

Recently it became possible to make individual crystals so large that one of them, or an aggregate of a few, is by itself large enough to serve as a sample for magnetic testing. The I-vs.-H relation for an individual crystal is very different from the relation for a mass of tiny crystals of the same material. In fact, the vector **I** is usually not parallel to the vector **H**. A magnetic field, applied to an ellipsoid cut from a single crystal, magnetizes it askew unless the field, and an axis of the ellipsoid, and an axis of the crystal happen to be all parallel to one another. In the polycrystalline mass, these deviations between direction of field and direction of magnetization must be averaged, and cancel one another out; for otherwise, the universally made assumption that I is parallel to H would not have been effectual. No doubt it is fortunate that Nature, with a rare benevolence, simplified the data first presented to the students of magnetism by this averaging and this cancellation. We cannot however conclude with safety that an assemblage of small crystals will behave just like an assemblage of equally many large ones. Evidently the size of the crystal must influence its I-vs.-H relations, or else the boundaries between adjacent crystals affect the magnetization, or there is something inherent which changes concurrently with the degree of crystallization. At any rate, whenever the crystallization of a sample is varied, the I-vs.-H curve is liable to feel it.

Composition and strain, temperature and current, state of crystallization—one must be prepared to find a new way of dependence of I upon H for each combination and every gradation of these; and yet the half has not been told. Whenever stress or heat are applied to a magnetizable substance, they alter its I-vs.-H relation, *not merely while they are being applied, but after they are withdrawn.* After such an experiment one may restore the original temperature and the original stress or freedom from stress, but the material is no longer quite the same. Vibrations and concussions, compressions and tensions and twistings, bending and tapping and cold-rolling and hammering, heating and cooling, annealing and quenching, *the very act of magnetization itself*—each of these is liable not merely to affect the I-vs.-H curve while it prevails, but to transform the substance permanently into another and a distinct ferromagnetic substance, with a system of magnetization-curves distinct from what the sample showed beforehand.

If we could see into the penetralia of a piece of iron, and discern the conditions and the arrangements of its atoms, it is probable that we should see that every such agency leaves behind it some definite and enduring change; and then we should not wonder at (for example) the fact that an iron wire, which undergoes the experience of being violently pulled and then relaxed, displays very different *I*-vs.-*H* curves before and after this adventure. In certain cases we do observe some sort of an attendant change, as for instance when the iron wire has been stretched so forcefully that it is permanently lengthened, or cold-rolled so vigorously that the X-ray diffraction-pattern due to its little crystals is affected. In other cases we observe no concurrent change whatever, and are forced to assume that there has been an internal alteration of the metal, for which there is no evidence beyond the testimony of the changed magnetization-curve. In the same way, we are prone to assume that when "the burnt child dreads the fire," something is altered within his brain-cells, for which there is no evidence except his change of conduct. As a rule, one would not speak of the brain-cells; one would say that the child has a memory of the painful burn. The ferromagnetic substance also changes its conduct after each experience, as though it remembered. No one, I presume, supposes that it actually has a consciousness which remembers; but the actual responsible alteration, whatever it may be, is often as far beyond detection as the alteration in the brain-cells. The resulting change in conduct, the result of this "memory" of the metal, is what is known as *hysteresis*.

All this makes the designing of a model for a ferromagnetic substance a very difficult and perplexing problem indeed, as we shall discover in due time. For the moment we are concerned only with knowing how much of the biography of a piece of metal must be recorded, in order to give background and value to a determination of its *I*-vs.-*H* curve. A curve inscribed "*This is the I-vs.-H curve for iron*" would not be worth much, no matter how carefully it had been determined nor how nearly pure the iron had been. At this point the physicist must betake himself to the foundry and the rolling-mill, and confer with the metallurgist, and learn the usage of a number of uncouth words such as *swaging* and *sintering* and *cold-working* and *quenching*, and grasp the distinction between cast-iron and wrought-iron and pig-iron and soft steel and hard steel, and observe a number of processes which were discovered so long ago that originally they were practiced without the least assistance from the guiding hand of pure science. The curve for his sample of metal must be labelled with the processes which the sample underwent before and after it came into his hands. Even yet it is not completely settled how many of the details of these processes should be recorded, nor how far back the history of the sample should be traced. One piece of knowledge, however, dispenses us from the risks of this uncertainty; it is known that a long-continued annealing,* followed by a gradual cooling, obliterates the traces of earlier experiences; and consequently a sample of unknown (or known) antecedents can be restored, by putting it through this process, to a standardized initial state.

Imagine, then, a sample of nickel which since its latest rejuvenation by annealing has undergone a recorded set of experiences; for instance, that it has been "stretched almost to the point of rupture, bent into a circle, and allowed to restraighten itself" (I quote an actual case investigated by R. Forrer). It might now be thought that, so long as the greatest care is exercised to avoid subjecting the metal to new stresses, concussions or heatings, the I-vs.-H curve would be fixed for good. Not so! for in order to determine the I-vs.-H curve, the metal must be magnetized; and magnetization, like stress and heating, is one of the events that leave an imprint, one of the experiences which the metal remembers. If two I-vs.-H curves are measured in succession, the second is generally not the same as the first; during the process of ascertaining the first, the material was changed into a new one. To predict or classify an I-vs.-H curve, one must not only know the composition of the substance, not only have records of its entire mechanical and thermal history since it was last rendered forgetful of its past by annealing, but also have the protocol of all its magnetizings since the last occasion when it was "completely demagnetized"—whether by the annealing which effaced all the memories, or by the gentler process † prescribed by Ewing which cancels the imprints of past magnetizations without destroying those of past stresses and heatings.

To make some choice among this staggering mass of data, it is suitable to concentrate one's attention on two, or rather on one and a group, of the infinite multitude of curves. The first of the chosen curves is obtained by applying to a sample which is freshly demagnetized a magnetizing field H_e which is increased by consecutive small steps, and measuring the field of the magnet after, or (by the method of the loop) the increase of the induction in the magnet during, each of these steps. From either of these sets of data, after making the allowances and the reductions indicated in the first section of this article, one may determine the I-vs.-H curve for steadily-increasing magnetizing fields applied to a piece of metal initially demagnetized.‡

* I use the word "anneal" to denote a long-continued maintenance at a high temperature, irrespective of the rate of cooling thereafter.

† By applying an alternating magnetic field of which the amplitude is at first greater than any field which has been applied to the magnet, and thence diminishes gradually to zero. However, the effect of this process is not quite thoroughgoing.

‡ There is a risk that the increase in magnetization at a certain step may be so great that, when due allowance is made for the demagnetizing effect of the magnet upon itself, it will be found that H has actually decreased in spite of the increase of H_e.

This is what is sometimes called the normal magnetization curve, sometimes the initial curve; I will adopt the latter term.

At this point it is well to recall that most of the curves actually found in the literature are B-vs.-H curves, not I-vs.-H. Since in the right-hand member of the equation $B = H + 4\pi I$ the second term is usually enormously greater than the first, a B-vs.-H curve usually looks exactly like an I-vs.-H curve plotted on a smaller scale. At very high fieldstrengths, however, a B-vs.-H curve continues climbing upward with a constant slope while the corresponding I-vs. H curve runs parallel to the H-axis.

The Initial Curve

The form of the initial curve is peculiar and distinctive. Departing from the origin of the (I, H) coordinate-plane, it ascends, bends upward, passes through a point of inflection, bends over but never quite turns downward; it goes off towards a horizontal asymptote, toward a maximum or *saturation* value of magnetization. Nearly all initial curves display these features, the point of inflection and the saturation; but in all other details, in the lengths and curvatures of the arcs before and after the point of inflection, in the scale of the curve and of its parts, they differ very much from one substance to another, and are altered very much by mechanical and thermal treatments.

Well-annealed substances, iron and nickel and permalloy for instance, display curves which tempt the onlooker to divide them into three segments: a slowly-rising and eventually upward-bending arc starting from the origin, a relatively steep-climbing portion including the point of inflection, a final arc drawing itself close up to the asymptote. A good example is shown in Figure 1. The distinction is accentuated by the hysteresis-loops which originate from the various points of the curve. In the prevalent theories of magnetization, as we shall eventually find, these segments are supposed to result from different processes occurring inside the metal. I will therefore adopt this separation of the curve into three parts, warning the reader to remember that at best there is always something arbitrary in subdividing a continuous curve, and at worst there are substances in which the division into three segments becomes quite impossible to make.

The first segment, extending from the origin to what some call the instep of the curve, may be regarded as a parabolic arc so long as the field is rather low—for iron and nickel, inferior to about one gauss; and for these metals it is sensibly a straight line so long as the field is below say a tenth, or to be safe a hundredth of a gauss.

This rather nebulous statement might be made precise by expressing I as a power-series in H, after this fashion:

$$I = aH + bH^2 + cH^3 + \ldots,$$

and citing experimental values of the coefficients $a, b, c, \ldots$; from this a student equipped with a measuring-instrument could determine

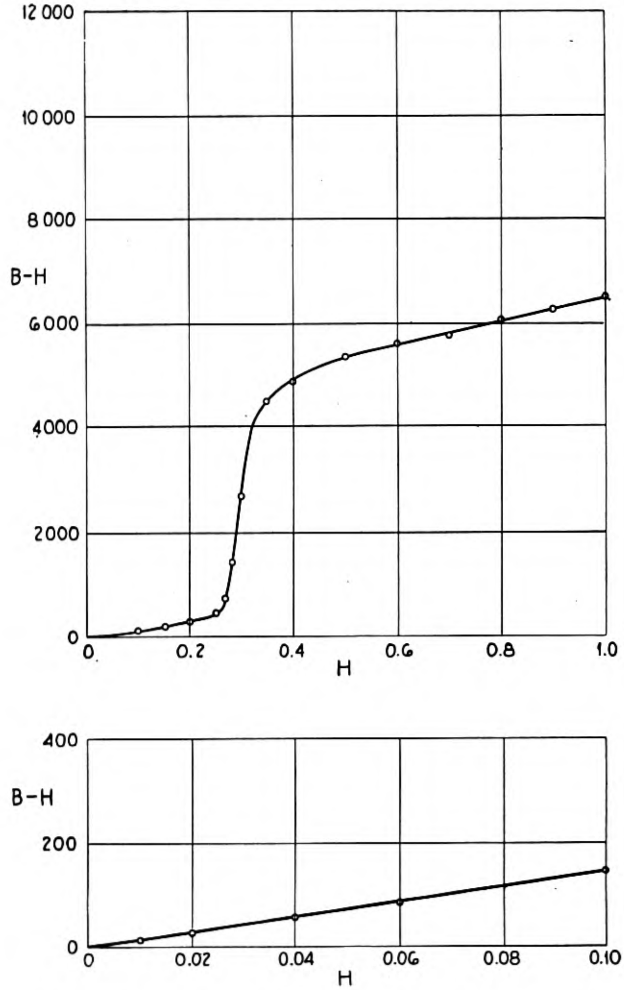


Fig. 1—Initial curve for a permalloy, displaying the three segments. The lower curve is part of the first segment of the upper, on a larger scale. (After L. W. McKeehan.)

the value of H at which the terms after the first become so small that with his apparatus he could not detect them. I mention this to show how valueless is the mere statement that "the early portion of the magnetization-curve is very nearly straight."

Values of the coefficient a, which is frequently called *initial susceptibility*—the corresponding value of $(1 + 4\pi a)$ is called *initial permeability* and denoted by μ_0—are rather often determined; it is an important constant of each material. Some pairs of values of a and b are quoted by Ewing, others by Bidwell, and some others were determined by the pupils of Weiss. According to one of these latter (Renger) the values for very pure freshly-annealed iron at room-temperature are: $a = 49.9$, $b = 108$. Tempered steel however yielded values of 2.23 for a, and 0.032 for b; from which, and from a mass of other observations on metals hardened by stretching, one sees that the effect of hardening is to lower a a great deal and b a great deal more, so that the curve slopes less sharply upward and does not begin to bend appreciably for a much longer way. I cannot quote all of the relevant data; but it is worth remembering that Rayleigh made measurements so delicate that he was able to follow the curve (for unannealed iron) all the way from .04 to .00004 gauss. Over this range his magnetometer reported no variation in the ratio of I to H. For nickel the detectable upward curvature commences at a much higher fieldstrength—five gauss, according to Ewing.

The alloys of iron and nickel, containing more than 30 per cent of the latter element, develop extraordinary magnetic properties when they are submitted to certain heat-treatments,* as G. W. Elmen discovered towards 1915. For these, the first segment of the magnetization-curve shrinks to a small fraction of the length it has for iron; the two-term formula

$$I = aH + bH^2$$

becomes visibly inadequate at 0.02 gauss, as the curve sweeps upward into its rapidly-rising stage. The value of a for some of these "permalloys" is as great as 8000, the value of b as great as 4000.

As the value of H is increased the later terms in the power-series for I bulk larger, and eventually the first segment of the curve passes over into what I have called the second. In this second section the ratio of I to H rapidly rises, and attains the enormous values which form one of the distinguishing marks of ferromagnetic substances, and are responsible for much of their utility in the world of engineering. Plotted against H, the ratio of I to H appears as a curve with a high

* For samples of a certain specified shape and size, this is the heat treatment which was recommended: "They are first heated at about 900° C. for an hour and allowed to cool slowly, being protected from oxidation throughout these processes. They are then reheated to 600° C., quickly removed from the furnace, and laid upon a copper plate which is at room-temperature."

sharp maximum, and so also does the more-commonly-plotted ratio of B to H, the permeability μ.

$$\mu = B/H = (H + 4\pi I)/H = 1 + 4\pi(I/H).$$

Pure iron attains much higher values of permeability than does either of the other metals which can be ferromagnetic when pure. By careful purifying and long annealing, T. D. Yensen elevated $\mu_{max.}$

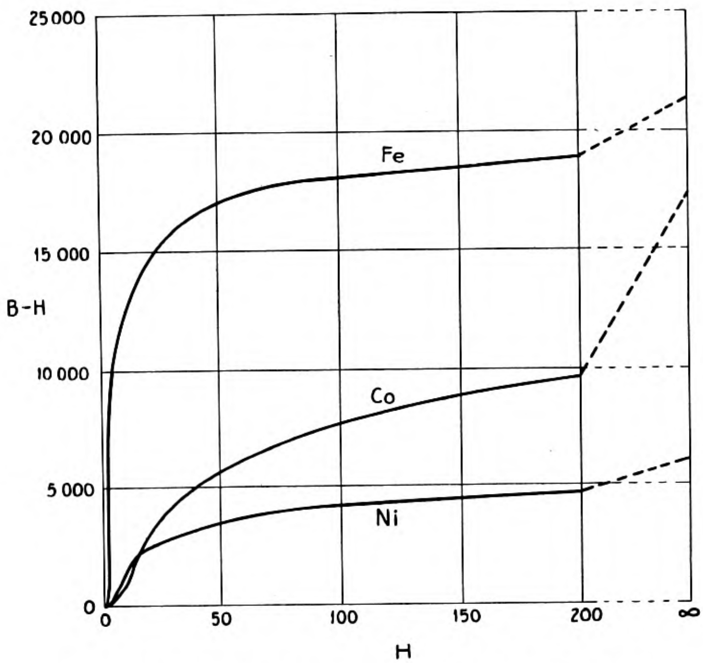


Fig. 2—Initial curves for annealed iron, nickel, and cobalt. (After L. W. McKeehan.)

for iron to 19000; the best recorded values for nickel and for cobalt are very considerably lower. Certain alloys of iron, however, leave the pure metal far in the rear; by slight admixtures of silicon (between 0.15 per cent and 4 per cent) Yensen produced materials which, after being melted in vacuo, annealed at a high temperature and very slowly cooled, developed a value of $\mu_{max.}$ as high as 66500. These in their turn were surpassed by the permalloys of G. W. Elmen for which $\mu_{max.}$ ascended past 100000.

Other alloys of iron, and in fact nearly all of them, are much inferior to the pure metal in respect of $\mu_{max.}$. Carbon in particular is pernicious; one per cent of this element mixed with iron brings the maximum permeability down to 350. A few per cent of manganese mixed into iron reduces I/H to the nearly-constant value of .03. Tempering, cold-working, forging, and drawing all tend to reduce the permeability. Since these processes render the metal harder in the literal sense of the word, the change which they imprint upon the magnetization-curve is called by association of ideas a "magnetic hardening." As $\mu_{max.}$ is reduced by any of these processes, the contrast between the three segments of the I-vs.-H curve diminishes, and in some cases there is scarcely more than the point of inflection left to mark the passage from the initial to the final range of the curve.

The final approach to saturation conforms to the law

$$I = I_{max.}\left(1 - \frac{c}{H}\right).$$

The value of the constant c is large for magnetically-hard materials, and small for the well-annealed samples for which the tripartite division of the I-vs.-H curve is obvious. In iron (I quote Weiss) saturation is approached within one promille at a field strength of 5500, in nickel at $H = 10000$. In permalloy it must be approached as closely with a field of a few dozens of gauss.

The saturation-intensity of magnetization, or *saturation* for short, is much more nearly independent of the present hardness and the past mechanical and thermal treatments of the material than the other features of the initial curve—much more nearly, therefore, a function of the chemical composition exclusively, than is any other single nameable magnetic quality. For this reason it is possible to present such a Table as the accompanying one with comparatively few qualifications. The first column of figures contains values of $I_{max.}$ obtained near room-temperature; the second, values measured at the temperature of boiling hydrogen, inserted here for future reference.*

* These values may be described as the "saturation magnetization of a cubic centimetre" of the materials in question. Dividing each by the density ρ of the material, we obtain the "saturation magnetization per gramme." Multiplying this by A, the atomic weight of the element or molecular weight of the compound (if the material is of either sort) we get the "saturation magnetization per gramme-molecule." This last is the quantity most often tabulated, being sometimes expressed in "magnetons" (units equal to 1126 C. G. S. units; cf. page 353). It may be advisable to recall that an *isolated* cube containing one cubic centimetre, or one gramme, or one gramme-molecule of material would not acquire the magnetization in question at any finite field, since it could not be magnetized uniformly.

TABLE

	$I_{max.}$ (20° C.)	$I_{max.}$ (20° K.)
Iron	1706	1742
Nickel	479	505
Cobalt	1412	
Alloy Fe_2Co	1880	
Permalloy Ni 78.5 per cent, Fe 21.5 per cent	870	
Heusler alloy Cu 75 per cent, Mn 14 per cent, Al 10 per cent	222	
Magnetite	490	518
Pyrrhotine	62	

The dependence of the initial curve upon temperature and strain is great and important; but it is expedient to reserve discussion of these variations to later sections.

The Hysteresis-Loops

Any ferromagnetic material has an infinite variety of hysteresis-loops, almost any one of which may turn up in practice; but I will limit this discussion to those obtained by a particular procedure, thus: Commence by demagnetizing the sample—increase H gradually to any desired value, denote this by H_0—decrease H gradually to and through zero, reversing it and bringing it to the equal and oppositely-directed value $(-H_0)$—return gradually to $+H_0$—return to $(-H_0)$—and so over and over again, ten or twenty times at the least. The point representing I as function of H, or B as function of H, traces out at first an arc of the initial curve extending as far as H_0; thenceforward it travels in sweeping detours passing around and around the origin, successive ones becoming more and more closely alike, until at last it settles into a routine of tracing the same oddly-shaped loop over and over again. I have spoken of the "memory" of the magnetic material; this process recalls the consolidation of memory into habit. The final habitual loop thus attained is the particular and chief *hysteresis-loop associated with* H_0. Demagnetizing the sample afresh and repeating the process with a new value of H_0, one gets another loop *; and in this way a family of hysteresis-loops can be determined, one for every point along the initial curve.

So long as H_0 is so low that the initial curve does not depart appreciably from a straight line, the hysteresis also is inappreciable; the point representing I (or B) as function of H goes back and forth

* The demagnetization may be dispensed with, if the new value of H_0 is greater than the prior one.

through the origin along the line of slope a (or μ_0). For this reason, the sensibly-linear part of the initial curve is often called the "reversible part." When it passes over into the perceptibly-upward-turning part, the hysteresis-loop becomes perceptible. Over a certain range its area varies as the cube of H_0, and Weiss gives this formula, in which the coefficient b is used with the same meaning as heretofore:

$$\textit{Area of hysteresis-loop} = \int HdI = \tfrac{3}{4}bH_0^3.$$

In the second segment of the initial curve, the loop swells out to its fullest amplitude. This forms one of the reasons for the division of that curve into three parts; the middle one is sometimes called the "irreversible portion" of the curve. There is no formula available in this region, except the oddly though not universally effective one discovered by Steinmetz, in which the area of the loop is related not to H_0 but to the maximum value B_0 attained in the cycle. This "law of Steinmetz" reads *

$$\textit{area of loop} = \eta B_0^{1.6}.$$

Values of the constant η are frequently quoted in describing magnetic materials.

When H_0 is carried far into the third stage of the initial curve, so that in each cycle I approaches within a few per cent of $I_{max.}$, the hysteresis-curve assumes the form of a wide loop prolonged at its northeast and southwest corners (I use the analogy of a map) by long slender projections which narrow down into mere lines. So long as I is nearly equal to $I_{max.}$, the point tracing the I-vs.-H curve passes back and forth along nearly the same path. The final stage of the initial curve is therefore also called "reversible." The Steinmetz formula here becomes invalid.

The reason for laying so much stress on the areas is well known. When a piece of magnetizable metal is carried through a cycle of magnetization, for instance by varying the current through an encircling solenoid in a cyclic manner, the battery supplying the current is found to have expended an amount of energy $\int HdI$ per unit volume; and the metal is found to be warmed to a degree indicating that an equal amount of heat energy has appeared within it.

* The formula of Steinmetz is more general; it applies to hysteresis-loops executed between any two (not overly great) values of induction B_1 and B_2, and for these assumes the form

$$\textit{area of loop} = \eta \left(\frac{B_1 - B_2}{2}\right)^{1.6}.$$

It is clear that B_1 and B_2 must be given opposite signs if directed in opposite senses.

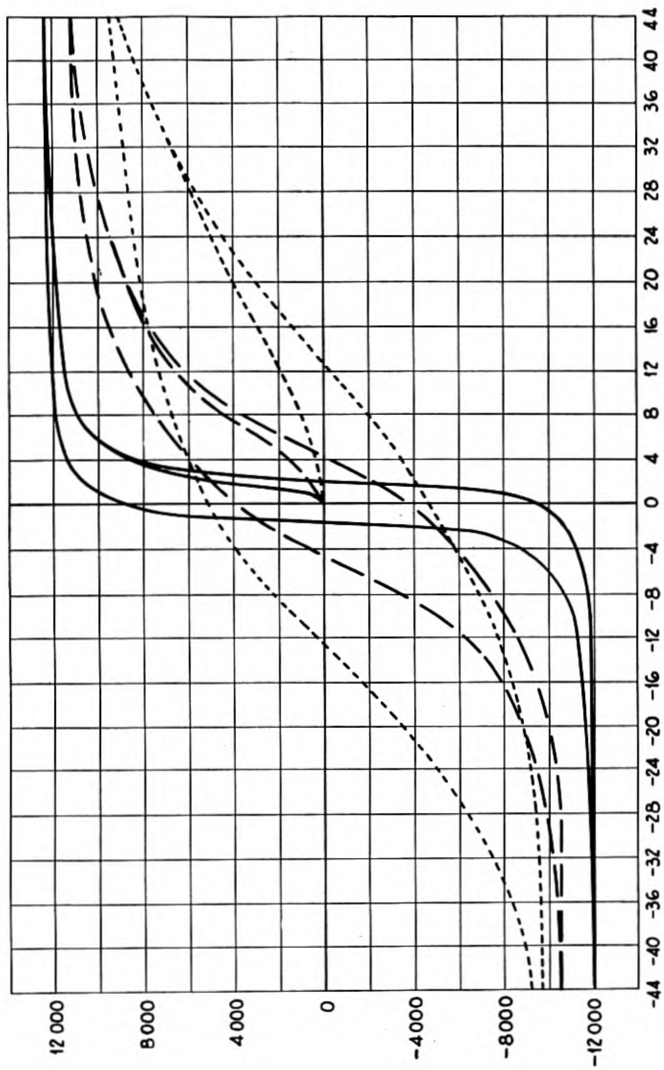


Fig. 3—Initial curves and hysteresis loops for annealed iron (continuous), harddrawn iron (dotted), and hard steel (dashed). (After F. Auerbach.)

We do not know how the transformation from electrical energy to thermal energy was effected; but we do know that so much electrical energy vanishes, and that so much thermal energy appears. A broad hysteresis-loop therefore signifies that there will be much dissipation of energy if the sample is exposed to cyclic magnetizing forces, as usually happens in electric machinery in which magnets play an important rôle; and the heat developed is not merely a sign of energy gone to waste, it is often detrimental to the material, and a bad contribution to that unforgettable history which the magnet is forever piling up. For these reasons the discovery of a new ferromagnetic material of low hysteresis is always welcome.

As a rule, narrow hysteresis-loops go with high values of initial permeability, and with initial curves easily divided into three stages, and with early saturation; and wide loops go with initial curves which rise slowly and bend upward slowly and display no sharply-marked second segment and approach very tardily to the saturation value. Magnetic hardening, in the sense which I earlier defined, accentuates hysteresis; and the agencies which bring it about—tempering and mechanical hardening and the admixture of certain elements in small quantities, such as carbon into iron—widen the loops and augment the generation of heat. These effects are frequently described by giving measurements of the heat W generated in a single cycle of magnetization in which B is carried back and forth between standard values $+B_0$ and $-B_0$ of the induction-measurements which in their turn are cited by giving the corresponding values of $W/B_0^{1.6}$, the quantity η of the formula of Steinmetz.* The value of this quantity is only .00032 for very pure well-annealed iron, leaps to 0.015 when one per cent of carbon is added, leaps again to 0.034 when the so-constituted steel is tempered; while the addition of silicon to iron, the very process which raises $\mu_{max.}$ to values excelled only by permalloys, brings the value of η down to .00011. The permalloys themselves are still more eminent in this regard, some of them having hysteresis-loops only a sixteenth as great in area as those of pure annealed iron.

To give the area of a loop is not always sufficient; its shape and orientation are very important for theory and for practice. The agencies which harden a material not only widen its hysteresis-loops, but rotate them clockwise around the origin, as the figures show. This rotation tends to decrease the intercept of each loop upon the

* I should emphasize that for many materials the "law of Steinmetz" is not accurate, so that strictly one should plot the actual curve of hysteresis-loss-*vs*-B_0, instead of making a single measurement and using it to determine η by the assumption that the "law" is valid.

axis of I or B, and increase the intercept upon the axis of H.* The former of these intercepts, representing as it does the magnetization which the metal retains when the external field has been reduced to zero, is known under the names of *residual magnetization* and *remanence* and *retentiveness*. The last two of these words, and also *residual magnetism*, are used in a general sense, to denote the property of not losing magnetization altogether when the magnetizing field is withdrawn. The intercept on the axis of H, representing as it does the force which must be applied oppositely to the direction of the prior magnetization in order to annul it altogether, is known under the names of *coercive force* and *coercivity*.

Residual magnetism was the first of magnetic phenomena to come under human notice. If the pieces of magnetite (lodestone) in the fields of Asia Minor had not been able to retain the magnetization which they had acquired in past ages, the Greeks would never have observed nor produced a magnetized metal; if steel needles rubbed against pieces of magnetite or held parallel to the earth's field and "smartly tapped," as the English textbooks say, could not retain the magnetic moment they so acquire, there would have been no compass-needles; the discovery of magnetism would probably have waited upon that of electric currents. Residual magnetism is the property to which the intercept of the hysteresis-loops gives a definite and definable meaning.

The greatest remanence, usually called *retentivity*, is attained after the material is magnetized to saturation. It may be as much as three-quarters of $I_{max.}$, or more.† Occasionally one finds samples of materials for which the ratio of the greatest remanence to $I_{max.}$ lies close to some simple fraction—in nickel, for instance, to 1/2. The greatest *ratio* of remanence to previously-attained magnetization, however, is obtained by choosing H_0 somewhere in the second segment of the pristine curve. In fact, there may be a long range of the second segment over which the difference $(I_1 - I_2)$ between the ordinates of the initial curve for any values H_1 and H_2 of the magnetizing field is practically equal to the difference between the values of the remanence in the hysteresis-loops for which $H_0 = H_1$ and $H_0 = H_2$ respectively. In other words: along the second segment of the curve, whatever added magnetization is given to the metal by increasing

* This is not a universal rule; samples of electrolytic iron studies by E. Gumlich and W. Steinhaus (*E. T. Z.*, 36, pp. 675–677, 691–694; 1915) which had been annealed at constant temperatures and cooled at various rates displayed intercepts on the I-axis which were much lower when the cooling had been rapid than when it had been slow; but the intercepts on the H-axis remained nearly unchanged.

† Ewing records an instance of remanence 0.96 as great as prior magnetization (hardened nickel under strong compression).

the field is *kept* almost intact when the field is annulled. Near the beginning and near the end of the curve, the magnetization which is conferred upon the metal by the field departs with the field. Ewing's theory of magnetization is strengthened by this fact.

The greatest remanence, as I have intimated, occurs with magnetically-soft materials. Magnetic hardening tends to augment the coercive force at the expense of the remanence. It does not follow that in constructing a good strong permanent magnet one should take a piece of well-annealed iron or permalloy. A substance of low coercive force is liable to lose its magnetization not only when it is exposed to a weak counteracting field, but also when it is bumped or jarred. A magnet which ceases to be one when dropped on the ground is not of much use in the compass or the automobile.* Great coercive force is much sought after in designing permanent magnets, and the alloys developed for this purpose are at the opposite pole of the ferromagnetic world from the permalloys. The maximum coercive force is attained after the material is magnetized to saturation; for it the name *coercivity* is used and should be reserved. The coercivity of iron, which when the metal is very pure and well annealed may be as low as 0.5 gauss, is elevated past 50 by alloying with one per cent of carbon, past 60 by a few per cent of tungsten, past 80 by a few per cent of molybdenum, up to 370 by amalgamating the iron with mercury. For the permalloys the values drop below 0.05. These figures naturally relate to samples already magnetized to saturation.

Other I-*vs.*-H *curves*

Still other I-vs.-H curves are obtained in special ways, a few of which I will mention.

If during the measuring of an initial curve the sample is continually shaken, or if after each change in magnetizing field it is traversed by a damped alternating current before the magnetization is read, the I-vs.-H curve rises very swiftly from the origin; it seems as if the first segment had been suppressed, the second rendered steeper than for the undisturbed sample. Fantastically high values of the ratio I/H are sometimes obtained in this way. Such curves are sometimes called "ideal curves," owing to an impression that they represent the true law of magnetization undisguised by accidental (?) influences.

If in the process of measuring a hysteresis-loop the observer stops

* There is an additional reason for not making permanent magnets out of substances of low coercivity. Suppose an ellipsoid of such a substance magnetized to saturation by an external field H_e; let H_e be reduced gradually to zero; the field $H = H_e - H_i = H_e - NI$ passes through zero long before H_e does, and when H_e finally falls to zero the value of I has fallen far below the true remanence unless the I-vs.-H curve runs nearly parallel to the axis of H.

short after the first reduction in magnetizing field following the attainment of the maximum applied field (the one which I above called H_0), returns to H_0, and alternates the field several times between H_0 and the inferior value $H_0 - \Delta H$, he finds that the magnetization settles down to a routine of alternating between definite values I_0 and $I_0 - \Delta I$. The limiting value of the quotient $\Delta I/\Delta H$, for small values of ΔH, is known as *reversible susceptibility*. It is a function of I; that is to say, if we select any particular value of I we always get one and the same value of $\Delta I/\Delta H$ when we impart that value of I to the metal, whether by mounting to it along the initial curve or the "ideal" curve or coming to it along any hysteresis-loop.

If the hysteresis-loop is described very rapidly and continuously, it retains its shape surprisingly closely until the frequency is raised into the hundreds of thousands. The initial permeability is still more nearly unaffected by rapidity of variation of field, remaining sensibly unchanged until the range of radiofrequencies is reached and passed. In the range of light-frequencies, however, it is reduced to unity.*

Magnetostriction

"Magnetostriction" is the clumsy name given to the divers very inconspicuous strains in a magnetizable body, brought about by the process of magnetizing it. As they are exceedingly small—a variation of any linear dimension amounting to four parts in one hundred thousand would be ranked as a remarkably big one—and as magnetizable materials are usually investigated in the form of long thin rods, the change in the length of such a rod resulting from a magnetic field applied parallel to its axis ("Joule effect") is the only magnetostrictive change which is often mentioned. Changes in the dimension normal to the field do, however, take place; a rod which expands lengthwise in a longitudinal field will contract sidewise, and *vice versa*. It used to be thought that the change in length just compensates the change in thickness, so that the net change in volume would turn out to be nil; but this turned out too simple to be true. A wire exposed to a longitudinal field and traversed by an electric current will twist itself (the "Wiedemann effect"). This occurs because the impressed field and the circular field due to the current itself are compounded with one another into a resultant pointing slantwise to the axis, so that any particular "line of force" can be visualized as winding in a helix

* In certain materials there is said to be a "magnetic viscosity," because of which the magnetization continues to vary for an appreciable time after an alteration in field is made and ended. The observations upon this are much confused by eddy-currents, and the question is still under debate.

around the wire from top to bottom, like the frieze of the Vendome Column; the expansion (or contraction) of the material along this line of force requires the wire to twist.

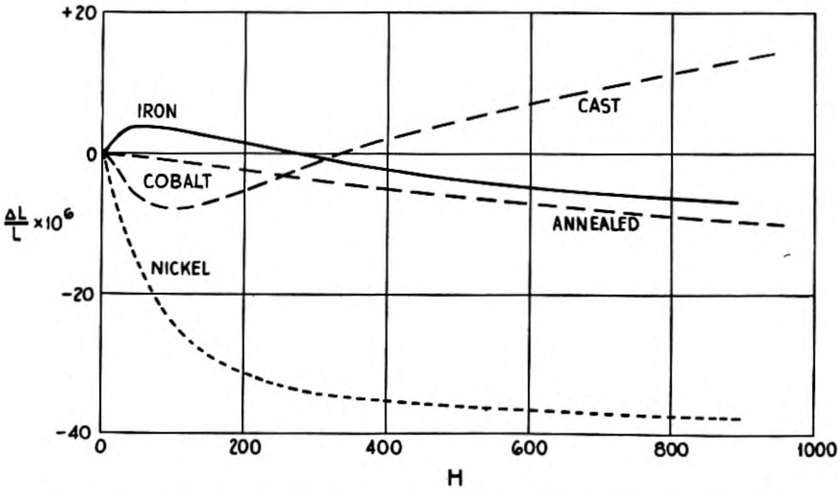


Fig. 4—Magnetostriction (Joule effect) in polycrystalline wires of iron, cobalt ("cast" and "annealed"), and nickel. (After K. Honda and S. Shimizu.)

It is customary to say that, in a gradually-increasing longitudinal magnetic field, nickel contracts continually; cobalt contracts at first, then returns to its original length, then expands; iron first expands, then returns to its original length, then contracts; the Heusler alloys expand continually. Unfortunately, some at least of these statements are valid only for samples which have been and are being treated in particular ways. One finds in the literature, for instance, the information that hard steel and very-well-annealed cobalt behave like nickel, shortening continually as the field is augmented. If the rules which I stated at first are really typical of the respective elements in standard states, then one may lay what emphasis he chooses on the fact that the four consecutive elements which are nickel, cobalt, iron and the manganese which is the essential element of the Heusler alloys are associated each with a different one of the four conceivable permutations of expansion and contraction.

The change in length, whichever its eventual sign, comes to an end when the material is magnetized to saturation. Intensity of magnetization is therefore the natural independent variable on which to consider magnetostriction as depending.

Quite the most exciting of the lately-discovered facts about magnetostriction is disclosed in Figure 4*a*, which consists of curves representing

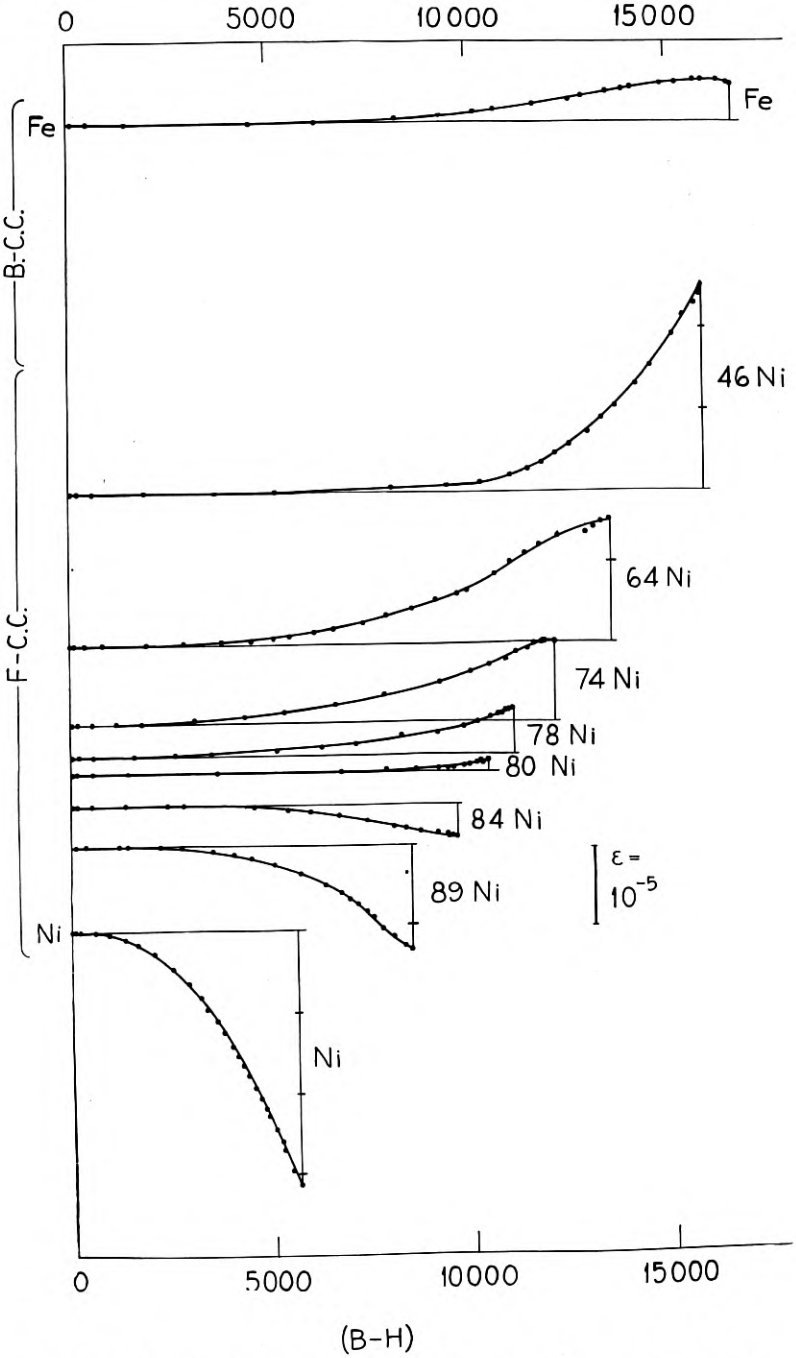


Fig. 4a—Magnetostriction in polycrystalline wires of annealed iron, nickel, and various permalloys. (After L. W. McKeehan and P. P. Cioffi.)

change-in-length for iron, nickel, and six permalloys in which the percentages of nickel are those indicated beside the curves. The term "permalloy," I recall, is applied to iron-nickel alloys containing more than 30 per cent of nickel, of which the initial permeability is remarkably high; the heat-treatments which these alloys had undergone conferred that quality on them, the nickel had been treated in quite and the iron in nearly the same way. The abscissa is intensity of magnetization, for the reason aforesaid; consequently the curves terminate when this reaches its saturation-value (not, however, attained in the experiments on iron and nickel).

These curves display the gradations from a steady lengthening reminding the onlooker of the initial lengthening of iron (not, however, followed by a contraction) to a steady contraction approaching the scale of that which nickel displays. Intermediate there lies an alloy which is influenced very little, indeed up to a high stage of magnetization it suffers no perceptible change at all; and this is precisely the alloy having the greatest permeability and the least hysteresis in the entire series. Upon this correlation McKeehan founded his theory of ferromagnetism.

This series of curves reveals other curious facts; for instance the extreme ineffectiveness of the first stages of magnetization in developing the strain—the 46 per cent-nickel alloy had expanded, by the time it was magnetized one third of the way to saturation, by less than one one-hundredth as much as it was destined to expand in acquiring the remaining two thirds of its final magnetization. This is significant; and more significant yet is the point, that when the magnetic field is applied to one of the permalloys containing less than 80 per cent of nickel and subject to a length-increasing longitudinal tension, the magnetostriction is much reduced—that is to say, the mechanical tension seems to have effected of itself a large part of that task of extension which else it would have been incumbent upon the magnetization to perform. It effects a great deal more, of course; the extension due to even a moderate load is vastly greater than the extension which even the greatest of magnetic fields could by itself ever cause; the point is, that the former extension seems to include the latter. Furthermore, elongation by tension is found to produce just as great and no greater an increase in the electrical resistance of a permalloy wire than the much smaller maximum elongation attending magnetization. Yet tension by itself does not magnetize; hence the change which it produces inside the wire does not entirely overlap the effect of magnetic field. It is also true, as one would expect, that tension acting upon a permalloy containing more than 80 per cent of nickel so affects it that the magnetostriction is increased—the tension seems as it were to have undone something, which the magnetic field must restore before proceeding to the contraction which it operates upon the unstrained metal.

Effect of Tension upon Magnetization

The effects of strain upon magnetization are very complicated, and one would almost despair of ever being able to interpret them, were there not certain relations between them and the effects known collectively as "magnetostriction"—between, to take the simplest instance, the influence of magnetizing upon length and the influence of lengthening upon magnetization—which indicate that law and order reign even in this seemingly chaotic field.

I mention the simplest instance only. A nickel wire, as we have seen, shortens when magnetized parallel to its length; well, when such a wire is shortened by compression, it becomes more magnetizable, the value of I and the value of I/H produced by a continually-applied field H increase; when it is lengthened by stretching (a much easier, consequently a much oftener performed process!), its susceptibility falls off greatly. An iron wire is lengthened when magnetized a little, shortened when magnetized strongly; when it is lengthened by stretching, the magnetization which a weak field imparts to it is increased, that which a strong field imparts to it is diminished; the magnetization-vs.-field curves for different extensions intersect one another somewhere upon the "second segment." Again, a cobalt wire, when lengthened by stretching, has a lower susceptibility in weak fields and a higher susceptibility in strong fields than it does when untensed; this corresponds to the rule governing the magnetostriction of cobalt. To the Wiedemann effect there correspond a magnetization which occurs when a wire carrying a current is twisted, and a rush of current which occurs on twisting a wire already magnetized. The signs of these effects, and of various others, vary from one ferromagnetic metal to another, and vary in iron and cobalt when the magnetization is sufficiently varied, in the ways which may be deduced from the corresponding magnetostrictive effects.

The variations in magnetization produced by extension may be very much more striking than the variations in length produced by magnetization. In nickel, for instance, the susceptibility of a wire may be reduced to a tenth of its pristine value by stretching the wire, although the utmost change in length which can be brought about by magnetization is less than one part in ten thousand.

The relations between the influence of magnetization on strain, and the influence of strain on magnetization, have been derived from the laws of thermodynamics. It appears that each of the several effects agrees with the theory insofar as the sign is concerned (for instance, tension applied to a wire which shortens when magnetized should diminish its susceptibility, and does) but not always in magnitude. I have not heard of anyone renouncing the laws of thermodynamics on this account.

Hysteresis plays a great part in the effect of tension on magnetization; if a constant magnetizing field is applied to a wire while the tension is being cyclically varied, the magnetization when plotted as function of extension follows a hysteresis-loop. Also the first application of a load to a wire is likely to make a sudden and violent change in the value of I. Some avoid these troubles, or try to, by shaking the wire continually or by continually applying an alternating magnetic field during the measurements. These introduce further complications. In fact, if all the data that could be assembled concerning the effect of strains upon magnetization were to be sought out, I suspect that "the world itself could not contain the books that should be written."

The Barkhausen Effect

Imagine once more a piece of some ferromagnetic substance, encircled by a magnetizing coil, through which the current is being steadily increased; encircled also by a loop, which is connected to the voltage terminals of an oscillograph, or to some other device which moment by moment records the electromotive force impressed upon the loop by the changing magnetization. This electromotive force, as I have said, is proportional to the rate-of-change dB/dt of the induction, for which the changing of the magnetizing field is responsible. It is a measure of the rate of magnetization of the sample girdled by the loop. The magnetizing field is being increased continuously; were the magnetization also to rise continuously towards its saturation-value, as we should probably expect, the voltage-curve would be smooth. However, when a sensitive oscillograph is used, the curve is a succession of sharp teeth. The magnetization of the sample evidently proceeds by small but sudden jumps. These can be shown—in the most literal sense of the word "to show"—by connecting a telephone-receiver through an audion-amplifier to the loop. Listening at the receiver, one hears a rustling or a crackling sound; it has been compared with rain beating upon a tin roof, also with coal rattling down a chute. Barkhausen discovered the effect in this way.

By increasing the magnetizing field very slowly it is possible to space the peaks in the oscillographic curve, or the clicks in the receiver, so widely that the bigger can be counted. Listening to the separated clicks, van der Pol estimated that the process of magnetizing a cubic centimetre of iron or of an iron-nickel alloy involves several thousand of the jumps. It is also possible to measure the area under each of the larger peaks in a curve obtained with a good oscillograph, and calculate from it the magnetic moment of a magnet, the sudden creation of which within the substance would have resulted in just such a peak. One observed by E. P. T. Tyndall could have come about through the sudden creation of a magnet of moment .0027. The word "creation" must not be taken too literally; it might imply, for instance, that two adjacent magnets were at first pointed contrariwise to one another, and one of them was suddenly wheeled around by the field, so that they ceased to neutralize each other. Data such as that just cited from Tyndall would then indicate the sizes of the magnets preexisting in the substance; data such as those of van der Pol, their number. Both sets of data show that one cannot identify these magnets with individual atoms; they are too large (the moment .0027 is as great as that of a piece of saturated iron 0.12 mm. on a side) and too few. Neither can they be identified with individual crystals; a piece composed of a single crystal makes as much noise in the receiver, while being magnetized, as a fine-grained sample. The data suggest that ferromagnetic metals are built up out of magnetic units larger than atoms and smaller than crystals—a suggestion which to the theorists is often extremely acceptable. It is also a welcome fact, that the peaks and the crackling are associated with the steeply-sloping segments of the magnetization-curve, while the initial and final nearly-horizontal arcs of the curve are smooth and silent.*

Magnetization of Single Crystals

Ferromagnetic crystals large enough to be studied are only just ceasing to be a rarity. Only two sorts occur in Nature: those of magnetite (a modification of one of the oxides of iron, Fe_3O_4) and those of pyrrhotine (a sulphide of iron, Fe_7S_8). To procure single crystals of a metal or an alloy, it used to be necessary to wait on the hazards of the foundry, out of which there might arise at long intervals a single large uniformly-crystallized lump. This condition prevails

* Attention must be drawn to the possibility that the peaks in the curve, or the clicks in the sound, are due to fortuitous coincidences of events individually too insignificant to be perceived. Should this turn out to be the case, the Barkhausen effect would resemble the Schroteffect of thermionic emission, and the interpretation of the data would be changed.

no longer; there are methods for producing large single crystals of metals at will, whether by direct solidification from the melt or by suitable treatment of the masses of randomly-disposed minute crystals which blocks of metals usually are; and there are methods for determining the orientations of the axes of these crystals by means of X-rays. So lately have these methods been developed (they are outgrowths of researches of the last ten or fifteen years) that the first data concerning the ferromagnetic crystals, except for some relating to magnetite and pyrrhotine and a very few early measurements on iron, are only now appearing. One has at times a feeling that these are the first really significant data, the only suitable foundation for a theory of ferromagnetism; that the properties of a polycrystalline rod or wire or ellipsoid do not form a proper basis for theorizing, not being even a simple average of the properties of single crystals oriented in all directions, but a deformed and distorted average infected by the crowding and the cramping and the squeezing which the little crystals perpetually inflict on one another.

All but two of the well-known ferromagnetic substances crystallize in the cubic system. (The exceptions are pyrrhotine and one modification of cobalt, which conform to the hexagonal system). In cubic crystals, directions parallel to the edges of the cubes, to their diagonals, to the diagonals of their faces, are called the tetragonal, trigonal, digonal axes, or the quaternary, ternary, binary axes respectively; the planes to which these directions are perpendicular are called (100) planes, (111) planes, (110) planes respectively. This is as much of the technical language of crystal analysis as we shall require. Of the three lattices in which atoms may be arranged in a cubic crystal—simple cubic, body centred, face centred—iron adopts the second, nickel and cobalt the third. The iron-nickel alloys containing more than 30 per cent of nickel copy the nickel lattice (the permalloys belong to this class) while those containing less than 30 per cent of nickel imitate iron. Many other metals which are not ferromagnetic have cubic lattices of the second or third type, none at all a lattice of the first; it is therefore futile to look for any correlation between ferromagnetism and the arrangement of the atoms.

When a magnetic field is applied to a crystal, it produces a magnetization which is not parallel to the acting field—to the resultant, I mean to say, of the applied field and that due to the "demagnetizing effect of the poles"—unless this resultant is parallel either to a tetragonal or to a trigonal or to a digonal axis. If we apply a field parallel to the axis of an ellipsoid or a long rod, cut from a single crystal in such a way that this axis is parallel to one of the specified directions, the *I*-vs.-*H* curve mounts much more swiftly than does the normal curve for polycrystalline iron; the first segment is very short, and the second passes into the third while the field is still low. The slope of the first part of the curve, that is to say the initial susceptibility, is greatest when the axis of the rod is a tetragonal axis of the crystal, less if it is a digonal, least if it is a trigonal axis; though the differences (Fig. 5) are not great. This is sometimes expressed by saying that iron is most easily magnetized along the tetragonal axis, less so along the digonal and least along the trigonal. Magnetization curves consisting of three or four straight lines meeting at sharp corners have been observed by two of the recent students of single crystals, but not by two others; I infer they are still debatable. The saturation value of *I*, whether it be

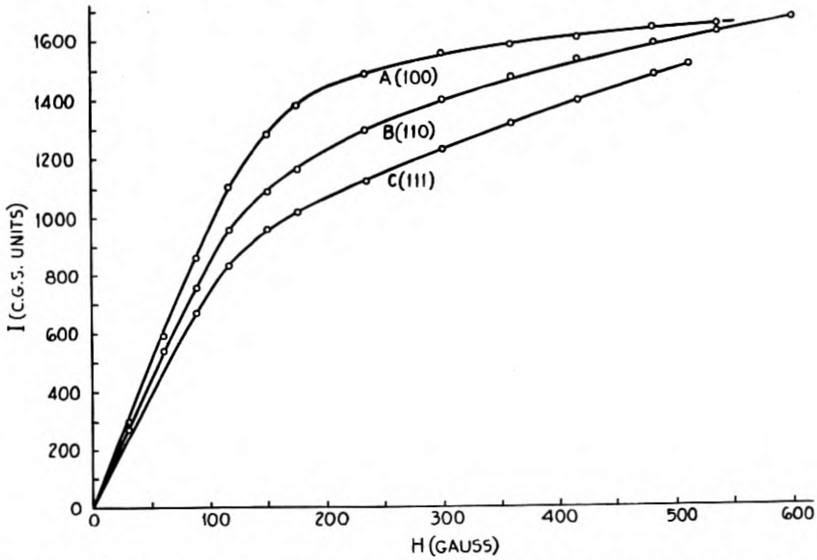


Fig. 5—Initial curves of a single crystal of iron, magnetized parallel to tetragonal (100), digonal (110), or trigonal (111) directions. (After W. L. Webster.)

attained soon or late, seems always to be about the same—another of the reasons for attaching a peculiar importance to it. Honda in fact obtained the value 1707, which he confronts with the 1706 given by Weiss for polycrystalline iron; but this is an agreement which looks too good to be true, or at least to be significant.

The hysteresis-loop for a single crystal is so exceedingly narrow that when it is plotted on any ordinary scale, its sides are too close to be distinguished. Measurements upon rods composed of many crystals, the average size of which varies from rod to rod, show that the area of the hysteresis-loop decreases quite steadily as this average size of the "grains" is diminished. This is a potent argument against all theories in which hysteresis is attributed to an arrangement of atoms in a uniform space lattice.

When a magnetic field is applied to an iron crystal in any direction not parallel to one of the axes, the magnetization is not quite parallel to the acting field. This manifests itself, for instance, when one cuts a disc out of a crystal and exposes it to a magnetic field in its own

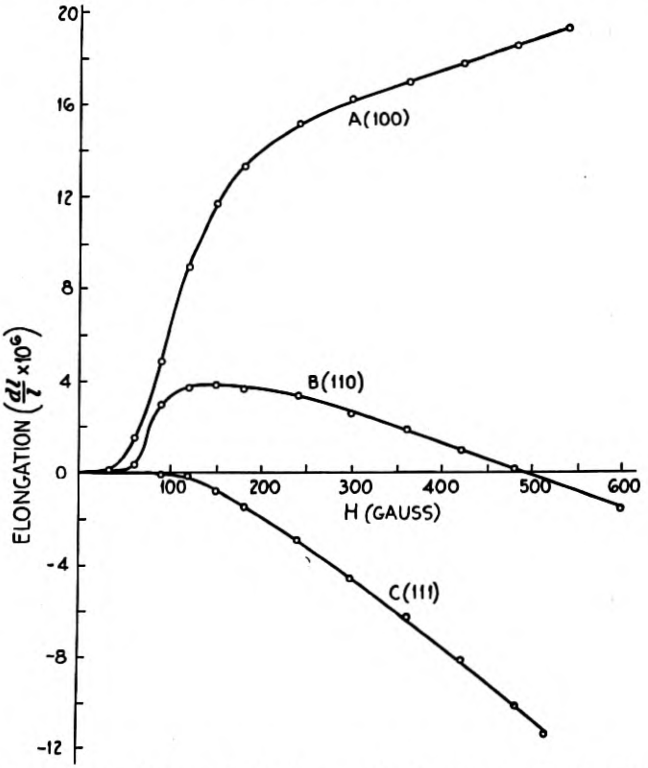


Fig. 6—Magnetostriction of a single crystal of iron, magnetized parallel to tetragonal, digonal, or trigonal axes. (After W. L. Webster.)

plane; it cannot rest in equilibrium until it has so turned itself that one of its three preferred directions lies parallel to the field, for otherwise there is a component of the magnetic moment which suffers a torque from the very field which evoked it. The angle between the vectors **I** and **H** seldom attains and never exceeds twelve degrees; when the field is kept constant in direction and varied in magnitude, this angle of deviation is less for very weak and less for strong fields than for some intermediate value of fieldstrength. In pyrrhotine, however, the angle may be enormous—a field inclined at no more than five or ten degrees to the hexagonal axis produces a magnetization which, when investigated by delicate methods, seems to lie exactly in the plane perpendicular to the hexagonal axis, which consequently is known as the "plane of easy magnetization." A sphere of pyrrhotine to which a bar magnet is brought up from the direction in which its hexagonal axis points does not seem to realize that the magnet is there, but if the approaching magnet is displaced a little

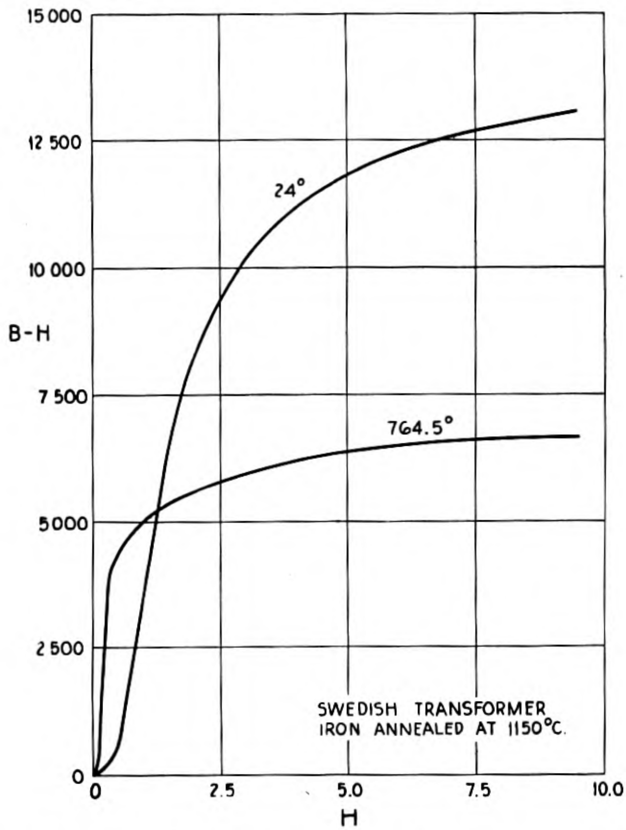


Fig. 7—Initial curves for Swedish transformer iron at two temperatures. (After D. K. Morris.)

sidewise the ball flies over to its surface at once. It will readily be seen what complications these facts introduce into the mathematics of predicting or describing the magnetization of an arbitrarily-shaped solid body—and in this connection it is well to remember that an ordinary polycrystalline mass of metal partakes as soon as it is strained, by pulling or rolling, of some of the properties of a single crystal.

Magnetostriction in single crystals has some very curious features. A crystal of iron unites in itself all the three modes of magnetostriction which have been supposed typical of iron, nickel and Heusler alloys respectively. A rod having a tetragonal axis along its length expands continually when exposed to a longitudinal magnetic field; a rod cut along the trigonal axis contracts continually; if cut along the digonal axis it first expands, then returns to its original length, finally contracts. The expansion in the first of these cases may attain twenty parts in a million—four or five times as great a value as one ever finds with a polycrystalline sample. This shows how great the extent to which the little crystals in an ordinary block of iron must interfere with one another when the block is magnetized.

Dependence of Magnetization on Temperature

As the temperature of a sample of iron is raised, its normal magnetization-curve varies in a manner suggesting the influence of tension;

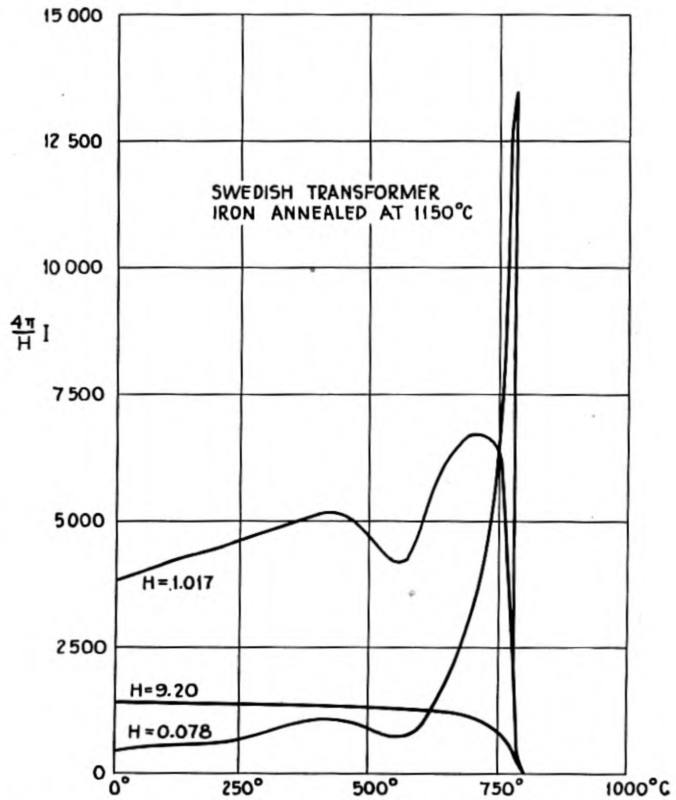


Fig. 8—Magnetization-vs.-temperature curves for Swedish transformer iron at three values of magnetizing field. (After D. K. Morris.)

the earlier part is exalted, the later part is depressed, so that the susceptibility increases in low fields and diminishes in high; curves obtained at different temperatures, not too far apart, intersect one another somewhere upon the "second segment" (Fig. 7). On plotting I or I/H for individual fieldstrengths as functions of temperature, one obtains curves which for very low fieldstrengths, such as 0.3 gauss for instance, are remarkably shaped (Fig. 8). The initial susceptibility rises to an enormous height at a temperature slightly above 700° C., and then precipitately falls almost to nothing—it does not quite vanish, but instruments of a much higher order of sensi-

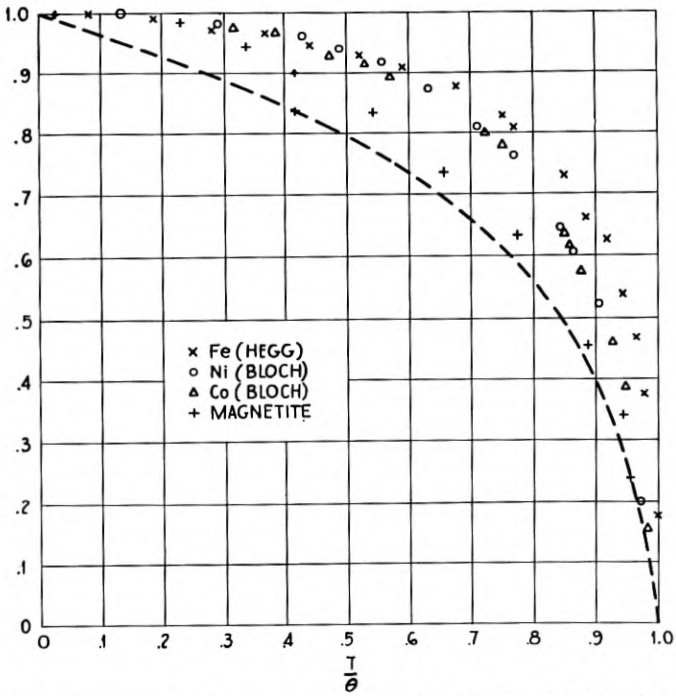


Fig. 9—Saturation-vs.-temperature data for iron, cobalt, nickel, and magnetite below their respective Curie-points, with a theoretical curve. Abscissa is ratio of absolute temperature to Curie-point temperature for each substance individually. (After P. Weiss.)

tiveness are required to detect or measure it beyond say 770°. At a somewhat higher fieldstrength, about 4 gauss, the I-vs.-T curve is nearly horizontal for a long way, and then declines gradually to the axis of H, which it reaches near 770°. At higher fieldstrengths the decline sets in progressively earlier (Fig. 8). At very high fields one obtains what is substantially the curve of $I_{max.}$ versus T (here the analogy with the effect of tension breaks down) which is shown in Figure 9.

The temperature at which I-vs.-T curves intersect the axis of T, or would intersect that axis were it not that they turn aside shortly before reaching it, is known as the *Curie-point*. For iron, values of Curie-point ranging from 768° to 790° are given; the differences seem to be due partly to uncertainties in deciding just where the I-vs.-T curves "would intersect the T-axis if they continued on downward without turning," partly to the indubitable fact that these intersection-points are not the same for I-vs.-T curves for different values of H, and partly to the use of other definitions of the Curie-point. For nickel, cobalt, and magnetite the Curie-points are in the neighborhood of 360°, 1130° and 550° respectively; and values are recorded for a considerable number of alloys.

The Curie-point is not the sign of what is properly designated as a "change of phase." Iron suffers changes of phase at temperatures near 900° and near 1400°, changes in which the atom-lattice goes over into an entirely different type, and a number of physical properties are sharply altered; but the Curie-point is not one of these, it is the locality of merely a rapid (though not absolutely sudden) change in magnetic properties and an evidently-correlated anomaly in specific heat.* As for the real changes of phase, they normally occur at temperatures so high that they do not influence the magnetization of iron below the Curie-point. Yet it is possible to bring one of the high-temperature modifications suddenly down into the low-temperature range, and then its magnetic properties are quite different from those of "ordinary" iron. In certain alloys this possibility is easy to realize; I will quote only the notorious case of a "nickel-steel" discovered by J. Hopkinson, which at 580° C. is merely one of the many non-ferromagnetic metals, remains so as it is cooled all the way down to zero, then turns suddenly into a modification which is strongly magnetizable and retains this state as it is being heated all the way back to 580° C. But indeed ferromagnetism of alloys is entangled with all the infinite complexities of the behavior and the internal changes of these complicated substances, and varies with all the variations of the more or less durable equilibria between their components.

Definition of Ferromagnetism

Ferromagnetism has sometimes been defined as "the kind of magnetism which iron exhibits"—an easy evasion, to which one is

* Contrary statements about iron are to be found in the early literature; but they are due partly to inaccurate experiments, and partly to the fact that the change-of-phase which in pure iron lies well above the Curie-point descends when carbon is progressively added to the iron, and before long comes into coincidence with the Curie-point; and if still more carbon is added, the "vanishing of ferromagnetism" takes place at the transition temperature.

sorely tempted to have recourse after the first few efforts to devise a better definition. Let us, nevertheless, at least take notice of a few of the alternative proposals.

Materials are classified into diamagnetic and paramagnetic and ferromagnetic. To distinguish those of the first sort is relatively easy, since in any of them the magnetization $\boldsymbol{I}$ called forth by an applied field $\boldsymbol{H}$ is antiparallel to $\boldsymbol{H}$ (in isotropic materials, at least; in crystals the angle between $\boldsymbol{I}$ and $\boldsymbol{H}$ lies between 90° and 180°). In materials of the second or of the third sort, the vectors $\boldsymbol{I}$ and $\boldsymbol{H}$ are parallel and point in the same sense, or at least are inclined to one another at angles smaller than 90°—provided, that is to say, that the material was demagnetized before $\boldsymbol{H}$ was applied. To distinguish between paramagnetic and ferromagnetic bodies, therefore, we must seek some other criterion.

The magnetization of iron, nickel, cobalt, certain of their alloys with one another and with other metals, and the Heusler alloys, may attain values enormously greater than those which can be impressed upon other substances with the highest possible fieldstrengths. One might therefore select some intermediate value for I, and say that all substances for which I may surpass this critical value are ferromagnetic, all others paramagnetic (or diamagnetic). In practice this is usually convenient, because of the great contrast between the substances which I just listed and practically all others. Among the elementary metals apart from the iron-cobalt-nickel triad, one of the most magnetizable is platinum, which shares a column of the periodic table with that triad; yet its susceptibility is only $2 \cdot 10^{-5}$, and an applied fieldstrength of 20000 gauss would impart to it a magnetization of less than one unit, which is utterly negligible compared to those which are easily imprinted even upon the less magnetizable of the substances which I named. The contrast is therefore great enough to be the basis for a useful definition. Yet it must be regarded as accidental, that in practice we are nearly always confronted with extreme cases of one sort or the other. If we travel along the iron-manganese or the nickel-chromium series of alloys (to take but two instances), or if we follow pure iron through a sufficient range of rising temperatures, we find a continuous series of intermediate stages between one extreme and the other; and in principle it is necessary to take account of these.

The magnetizations of iron, cobalt, nickel, certain of their alloys and the Heusler alloys increase, when the applied field is continuously increased, in the curious ways which I described above, attaining maximum limiting-values at fieldstrengths well within the practicable range; while with nearly all other materials I is apparently proportional to H as far as the field can be carried. Here again there is a contrast so great that it can serve as the basis of a useful distinction. Yet all the intermediate stages between the two extremes are exhibited by iron at the various temperatures between 700° and 800° C. Furthermore, there is reason from theory (as we shall see) for supposing that the magnetization of any substance would cease to be proportional to H and would approach a limit, if we could force the field to high enough or the temperatures to low enough values. In fact, there is at least one of the substances conventionally called "paramagnetic" (it is gadolinium sulphate) for which I was found to approach a limit, when the applied fieldstrength was increased while the substance was maintained at the unprecedentedly low temperature of 1.°3 K. It is therefore evidently something of an accident that in practice we nearly always meet either with substances for which the ratio I/H is constant within the accuracy of measurement throughout the feasible range of the fieldstrengths, or else with substances for which that ratio varies greatly and unmistakably with the field.

Presence or absence of hysteresis is the third and last of the usual criteria. Iron and cobalt and nickel and some of their alloys and the Heusler alloys exhibit hysteresis-loops, and residual magnetism, and coercive force; and the normal magnetization curve must be distinguished from curves obtained by other procedures for varying the applied field, and one must bother with demagnetizings or else take account of the prior magnetization of whatever sample he is working with. Other substances are free from these complications. Here also it is probable that in iron all the measurable features of hysteresis dwindle off continuously to zero as the metal is heated. On the other hand, it appears that gadolinium sulphate, in spite of acquiring a curvature in its I-vs.-H curve at extremely low temperatures, does not acquire hysteresis and residual magnetism. Perhaps, then, it is better to take the presence of hysteresis rather than the inconstancy of the ratio I/H as the sign of that curious quality, whatever it may be, which makes iron notable among metals.

The general conclusion seems to be the same, as for many other classifications—that is to say: It is possible to draw distinctions between "ferromagnetic" and "paramagnetic" substances, valid for extreme cases of the two types, not sharply marked for intermediate cases; but it happens that for the time being the intermediate cases are in practice not conspicuous; and consequently the distinctions—any one of the three which I mentioned—are useful and worth the making.

C. Theories of Ferromagnetism

To devise a theory of ferromagnetism is not necessarily the same task as to make a theory of magnetism. In studying the properties of paramagnetic and those of diamagnetic bodies, one finds many indications that the ultimate atoms of the elements are magnets of definite and seldom-changing moments, or at least may profitably be so regarded. The theory of line-spectra reinforces this opinion, and it is confirmed by the observations of Gerlach and Stern upon the deflections undergone by free-flying streams of atoms traversing a strong magnetic field with a strong field-gradient.* Now, to say that atoms are magnets is scarcely tantamount to giving an explanation of magnetism. On the contrary, the problem is merely pushed a step further away, and must eventually be faced again and either be solved by explaining why atoms are magnets, or else be given up by conceding that magnetism is one of the fundamental properties of matter. Yet it is quite logical and sensible to aspire to construct a theory of *magnetization*—of the gradual magnetizing of a substance by an increasing applied field, of the shapes of the I-vs.-H curves, of hysteresis-loops—out of the assumption that the ultimate atoms are permanent magnets. To explain the gradual rise of an I-vs.-H curve by postulating atoms which are already magnetized to saturation, to explain hysteresis by postulating atoms which individually have no hysteresis—these would be triumphs not open to the objection made against many "explanations," that they are achieved by ascribing to the atoms the very properties to be explained.

We shall presently make the acquaintance of "elementary magnets"—hypothetical beings, of which each magnetizable substance is supposed to consist. To these we shall assign, for the time at least, definite and unchangeable magnetic moments. A magnetic field applied to an assemblage of such magnets could not change the moment of any. Yet it could change the net magnetic moment of the assemblage, which is the resultant of the moments of all the individuals; for it could, directly or indirectly, cause the elementary magnets to align themselves along its own direction. The assemblage, the substance, would be magnetized *not through magnetization of the individuals but through orientation of the individuals* which make it up.

That idea is an old one; but by itself it is nearly useless. We must think of some agency which could combat the tendency of the elementary magnets to align themselves along the field; for there must be such a one, as otherwise the weakest possible field would magnetize each substance to saturation; which is not the case. The most

* I refer for these to my *Introduction to Contemporary Physics*, pp. 48–50, 383–393.

celebrated theories of magnetization rest upon speculations about the nature of this agency which fights against the field.

On considering the unsurpassably simple system composed of *only two* elementary magnets close together, J. A. Ewing discovered that their interactions are such, that they can prevent each other from aligning themselves immediately along the field; one can almost say that they "interlock," and they interlock in such a way, that the pair of them displays a tripartite *I*-vs.-*H* curve, and the quality of hysteresis, though neither separately has any such properties. Systems comprising a dozen, a score, or a multitude of such magnets, arranged in chains or in a cubical array, may be devised to imitate actual initial curves and actual hysteresis loops with stunning accuracy (Fig. 11). Such close agreements need not be overstressed. The astonishing feature of Ewing's discovery is (I think) that although each individual magnet possesses neither the quality of gradual magnetization nor the quality of hysteresis, a pair of them put close together possesses both. So great a result is attained from so simple an apparatus, that it seems very unlikely that any radically different explanation of either quality will ever be put forth. Whatever may be added to Ewing's model, its central idea will probably never be supplanted.

P. Langevin, devising a theory for paramagnetism, supposed that the agency which combats the aligning influence of the field is the thermal agitation of the magnetic atoms. Contrary to one's first impression, this theory is not easily visualized; but it establishes a union between paramagnetism on the one hand, and the great general principles of thermodynamics and equipartition of energy on the other. In the form in which Langevin put it forth, it does not account for hysteresis.

P. Weiss supplemented Langevin's theory by supposing that the actual magnetizing field prevailing inside a magnetizable substance is not that sum of the applied field H_e and the "demagnetizing field" H_i which I defined in Section A, but a combination of this sum with another term depending on the magnetization. As I stressed in Section A, experience teaches us nothing about the value of the true field inside a magnet; Weiss' assumption was therefore a perfectly legitimate choice, to be justified (if at all) by its fruits. One of these is, that it accounts for the presence of hysteresis at low temperatures and its absence at high.

Ewing's Theory

Ewing conceived a piece of iron as an assemblage of tiny bar-magnets, each endowed with a fixed and constant magnetic moment, and wheeling about a pivot under the combined influence of the impressed magnetic field and the magnetic attractions and repulsions of its neighbors.

Imagine a chain of long slender bar-magnets end to end, the positive pole of each almost touching the negative pole of the next—that is the equilibrium position which they would naturally assume, so long as no external field affects them. By preference, build such a chain out of pivoted magnets; for Ewing's model enjoys the singular merit, that it can be made out of actual magnets and exhibited to the eye. Now there is a remarkable feature of this chain: if a magnetic field is applied to it in some oblique direction, then so long as the fieldstrength is quite small the individual magnets incline themselves toward it slightly, each setting itself at the same angle to the direction of the chain which was originally the common direction of them all; and when the fieldstrength is gradually increased the angle increases gradually, but only up to a certain point—for suddenly, at a critical moment, all the bar-magnets very suddenly capsize, and set themselves in nearly the direction of the field. I use the word *capsize* to invoke the too-familiar analogy of the upsetting boat. As weights are piled upon one side of a boat, it responds at first by tilting gradually sidewise and downward; to each slight increment of the load it accommodates itself by finding an equilibrium-slant a little farther over; but eventually there comes a moment when balance and compromise are no longer possible; the boat cannot find a position of equilibrium except by overturning, and this it does, suddenly and irrevocably. Such is the behavior of a chain of bar magnets; and this is the property which adapts it for representing the general shape of an initial magnetization-curve such as I showed in Fig. 1, with its first slowly-rising arc followed by the rapid uprush and the final slow adjustment to saturation.

The overturned boat will not right itself even when the load which upset it is removed; will the chain of bar-magnets be equally unforgiving? The analogy is not perfect, except in one very particular case: if the angle between the direction of the chain (defined as the direction in which the north poles of all the magnets originally pointed) and the direction of the field is 180°, the capsizing will result in a right-about-face of each magnet and a reversal of the so-defined direction of the chain, and this reversal will persist after the field is annulled.

Suppose however that there is a multitude of chains oriented at random, so that half of them are inclined at less than 90° and half at more than 90° to the direction of any strong field which we choose to apply. The field will cause all the bar-magnets to capsize (except those belonging to the few chains to which it is almost parallel);

and thereupon, those which belong to the chains originally inclined at more than 90° to the field are more than halfway turned around, and when the field is nullified they will realign themselves with their first associates, but every one will be reversed. Originally the net magnetization of the assemblage of chains was nil, for half neutralized the other half; now it is considerable, for half have been inverted. Its ratio to the total magnetization of all the chains when parallel is, in fact, one half. This consequently would be an adequate model for a substance of which the remanence is one half of the saturation-intensity.

Other values than one half for the ratio of remanence to saturation can be derived from Ewing's picture by choosing a suitable arrangement for the elementary magnets. Suppose, for another and a final example, that they are arranged in a cubic lattice, so that each has the choice (as it were) of orienting itself along any one of the directions parallel to the cube-edges. Chains of magnets may then form themselves along any one of six possible directions (counting the two opposite senses of any line parallel to a cube-edge as two distinct directions). In a demagnetized crystal, one may imagine that the elementary magnets in the lattice fall into groups or "complexes," within each of which all the chains are parallel, while from one complex to the next they change over from one to another of the six specified possible directions. In a demagnetized piece of metal composed of many small crystals oriented quite at random, there will be chains of magnets pointing in all directions. To such a piece of metal let a field be applied, increased to so great an amount that it saturates the material, and reduced gradually to zero. Whatever the direction of the field, it will be inclined at 45° or less to one or more of the six possible directions for the magnet-chains in every crystal. As the field is varied in the manner which I have described, the magnets in each crystal will be wheeled into parallelism, and subsequently will relapse into chains pointed in that direction (or those directions) which makes the least angle with the field. The ratio of remanence to saturation for a polycrystalline sample, resulting from this model, should then be 0.893.

By adjusting the disposable constants, Ewing's model may be made to predict not only the general shape of the I-vs.-H curve, but the values of fieldstrength and magnetization at which the first segment of the curve should pass into the second. Apparently no very pleasing agreements between experiment and theory have yet been attained in this way. Nevertheless I will show how the attempt is made; by doing so, I can at least bring out the influence of the various disposable constants upon the result.

The simplest form of Ewing's model * is composed of linear chains of elementary magnets. To analyze this it suffices to consider a system composed of two identical magnets, each so long and slender that it may be visualized as a pair of poles of equal polestrength M separated by the length L of the magnet, and both of them pivoted around their centres at points distant from one another by a spacing S which is only slightly greater than L (Fig. 10). If there is no external field, they come to an equilibrium, in which state both point in the same sense along the line of centres. If there is an applied field oblique to the line of centres, they come to an equilibrium in which both are deflected through equal angles from that line. Denote their angles of deflection by θ, the angle between the field and the line of centres by α, the fieldstrength by H, the distance between the adjacent unlike poles of the magnets by R. The distance R is equal to $(S - L)$ when θ is zero, and in general is given by the equation:

$$R^2 = L^2 + S^2 - 2LS \cos \theta. \qquad (1)$$

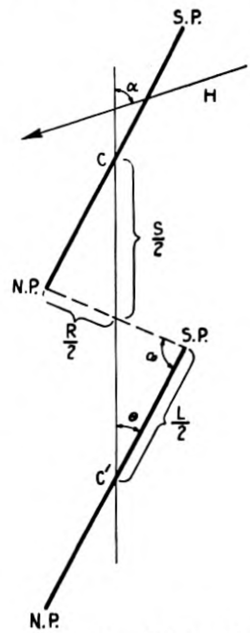


Fig. 10 — Illustrating the elementary magnet-pair of Ewing's theory.

The adjacent poles attract one another with forces M^2/R^2 directed along R, which result in torques T' upon each magnet:

$$T' = \frac{M^2L}{R^2} \sin \varphi = \frac{M^2LS \sin \theta}{2R^3}. \qquad (2)$$

The remote poles likewise exert forces upon the adjacent poles and upon one another, and torques upon the magnets; but it will be necessary to reduce these to relative insignificance by supposing the

* The next four pages resulted from an attempt to formulate what I take to be Ewing's objection to his own early model, which he phrases in these words: "Now it is known that in ordinary iron barely one per cent of the whole magnetism of saturation is acquired in the quasi-elastic stage before the effects of hysteresis set in. To conform to this condition the magnets of the model must have only a very narrow range of stable deflexion, and consequently they have to be set very near together with the result that in the old model their mutual control became excessive. A calculation of the force required to break up rows of pivoted magnets, of atomic dimensions, when set near enough together to satisfy the above condition, showed it to be many thousands of times greater than the force which is actually required, in iron to reach the steep part of the curve."

"clearance" $(S - L)$ between the adjacent ends of the magnets to be extremely small by comparison with S and L, and by considering only values of θ which are so small that R itself remains small by comparison with L; otherwise the equations will be hopelessly intricate, and they are more than bad enough even with this restriction. Happily the model possesses some of the required properties even when limited by this restriction.

The torque T exerted by the field H upon either magnet is given by

$$T = MHL \sin (\alpha - \theta). \tag{3}$$

The general condition for equilibrium is

$$T - T' = 0. \tag{4}$$

The special condition for "neutral" or "labile" equilibrium, i.e. for the state of incipient capsizal, is

$$d(T - T')/d\theta = 0. \tag{5}$$

The values of H and θ, obtained by solving (4) and (5) as simultaneous equations, are the fieldstrength just sufficient to produce capsizal and the angle of deflection attained just before the overturn; they are obtained as functions of the variable α, and of the constants M, L, S which are features of the model.

Solving these equations, however, is easier said than done; they prove surprisingly intractable. Only in one particular case is the solution easy: we must choose values of α so near to 90°, and suppose the clearance and consequently the deflections so small, that the cosine of $(\alpha - \theta)$ may be set equal to zero. In this case equation (5) is reduced to the form

$$dT'/d\theta = \text{const. } d(\sin \theta/R^3)/d\theta = 0, \tag{6}$$

which, if we write a for S/L, is found equivalent to

$$(a^2 + 1 - 2a \cos \theta) \cos \theta = 3a \sin^2 \theta. \tag{7}$$

Putting $a = 1 + \epsilon$—so that ϵ stands for the quantity $(S - L)/L$, which by hypothesis is small—and neglecting powers of ϵ higher than the second, we arrive at the equations:

$$\cos \theta_c = 1 - \tfrac{1}{4}\epsilon^2; \qquad \sin \theta_c = \epsilon/\sqrt{2} = (S - L)/L\sqrt{2}, \tag{8}$$

for the value θ_c of the deflection just at the verge of capsizal; and putting this into equation (4), we get

$$H_c = \frac{M}{3\sqrt{3}(S - L)^2}. \tag{9}$$

Equation (9) gives the fieldstrength H_c which effects capsizal of a pair of magnets initially transverse to the field, and having a clearance $(S - L)$ extremely small compared with their lengths L. For a chain of magnets the value given for H_c would need only to be doubled; for any number of chains lying in the plane normal to the field, that double value of H_c would remain valid. It is, we see, proportional directly to M and inversely to the square of the clearance.

Multiplying the expression given in (8) for $\sin \theta_c$ by ML, we get the component along the field-direction of the moment of any magnet belonging to such a pair or to such a chain. If there were N magnets grouped in pairs or chains in the plane normal to the applied field, the magnetization I of the entire assemblage, parallel to the field-direction, would be $NML \sin \theta$. The magnetization I_c at the verge of capsizal

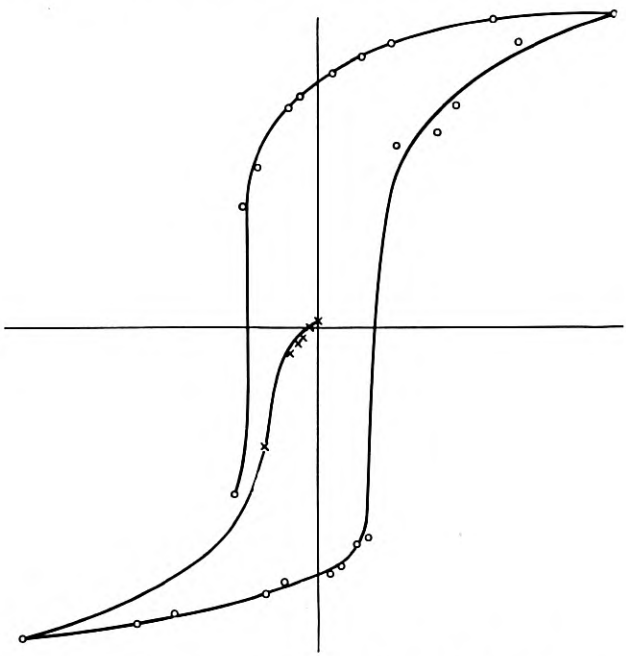


Fig. 11—Initial curve and hysteresis loop of an Ewing model composed of 24 pivoted magnets. (After F. A. Ewing.)

would be $NML \sin \theta_c$. The saturation-value of magnetization, I_{max}, would be NML. Thus we arrive at the equation for I_c:

$$I_c = I_{max} \frac{S - L}{L\sqrt{2}}, \qquad (10)$$

beside which we may place equation (9), reshaped and with allowance for the doubling:

$$H_c = \frac{2I_{\max}}{3\sqrt{3}NL(S-L)^2}. \tag{11}$$

It is now obvious that, in the case of a material for which $I_{\max}$ is known, we have apparently three disposable constants N, L, and $(S - L)$. However, the ratio of $(S - L)$ to L must be a very small fraction; otherwise the assumptions from which the equations were deduced would not be valid. This ratio is determined by the ratio $I_c/I_{\max}$. If we take for I_c the value of magnetization somewhere near the division between the first and second segments of the initial curve, we find $I_c/I_{\max} = .01$ for soft iron (I quote Ewing) or about .05 for the permalloy of which the curve is exhibited in Fig. 1. Now we have the ratio of $(S - L)$ to L fixed, and ostensibly two disposable constants left. However, if we assume that each elementary magnet is an atom and each atom an elementary magnet, both of these are determined by the crystal lattice of the metal. Nothing remains adjustable; a definite value is imposed by the theory upon H_c. This value is enormously too great.

It is clear that the situation could be saved by dropping the assumption that every atom is a magnet, so that the constants N and L might again become freely disposable. Ewing proposed another way of escape—a modification of the model involving the introduction of a fourth constant. He invented a system composed of three magnets with their centres in a line, the two outer ones fixed and pointing in opposite senses along the line of centres, the middle one free to revolve. The polestrengths of the outer magnets, M' and M'', are supposed to differ slightly; then, when no outside force is acting, the middle one comes to an equilibrium in which it points in the same sense as the stronger of its neighbors. When a field is applied in a direction inclined at α to the line of centres and steadily increased, capsizal occurs at a certain value of fieldstrength H and the corresponding value of deflection θ. When the clearances are small and α is very nearly 90°, the equation for θ is equation (8) with an unimportant change in numerical factor; while the equation for H is changed, in that $(M' - M'')$ now stands in the place of M. This is the new constant introduced into the model.

Ewing supposed that the pivoted magnet of his model might be the analogue of an internal electron-orbit of the iron atom, while the fixed neighbors might correspond to external distributions of whirling electrons, in the periphery of the same atom or in neighboring ones. This notion is endangered by the discovery that a single crystal of iron displays only a slight degree of hysteresis, much less than a polycrystalline mass—a discovery which likewise weakens the force of calculations of remanence based upon the assumption of a cubic lattice, such as I gave earlier. In fact, it seems quite probable that in the course of assimilating the newly-acquired data concerning single crystals, all of the numerical agreements hitherto derived from Ewing's and other theories of ferromagnetism may be swept away.

Nevertheless the basis of Ewing's theory is likely to persist; for it has two great advantages which are nearly independent of numerical agreements. Hysteresis is derived from an atom-model in which nothing of the nature of hysteresis is introduced by postulate; and the general effect of mechanical and electrical jerkings, bumpings and jogglings is explained in a way which seems most natural and plausible to our mechanical instincts. As for the first point: to explain hysteresis by the mutual interactions of magnets which in themselves are constant and do not possess it is so eminently satisfactory a solution that any theory or model in which hysteresis is introduced *ab initio* or derived from some extra assumption will start under a great handicap. As for the second: to take one illustration, it is well known that a demagnetized piece of iron exposed to a weak field, and endowed thereby with the moderate magnetization corresponding to some point or other on the first segment of the initial curve, becomes enormously more intensely magnetized when it is jolted or jerked. Visualized by Ewing's model, this seems the most natural thing in the world: the elementary magnets which were on or close to the verge of capsizing are pushed over that verge by the mechanical shock. Equally natural seem the annulment of the residual magnetism of a piece of iron, by mechanical shocks and jerks; the like effect of rapidly-alternating magnetic fields; the tendency of a current along an iron wire to favor magnetization of the wire; and the fact to which I alluded earlier, that in a very strong rotating magnetic field a piece of iron does not grow hot, and consequently there can be no hysteresis-loss. This last-named feature may be visualized by supposing that the chains, having been once completely broken up, do not form again as the magnets are whirled round and round. It seems natural also to expect that as the temperature is raised, the chains will be broken up by thermal agitation, and the reversible first segment of the initial curve will mount more sharply and continue longer. In trying to deal with the effect of temperature, however, we soon reach the limits to which Ewing's theory can be forced; and another method of attacking the problem of ferromagnetism recommends itself.

Weiss' Theory

There is another theory of magnetization, built upon an entirely different basis from Ewing's—a basis involving the notion and in fact the definition of temperature. To import temperature into theories of magnetism is clearly most desirable, considering how great is the influence of that variable upon the I-vs.-H curves; an influence so great, indeed, that when a sample of any ferromagnetic substance is made sufficiently hot, all the distinctive features of ferromagnetism depart from it. In developing Ewing's model, it is easy to say that as the temperature is raised the little magnets are more vigorously agitated, the bonds which are responsible for remanence and coercivity are more frequently ruptured; but such statements, though plausible, lack precision and hold out no promise of numerical agreements between theory and experience. That being the case, it seems unreasonable to expect numerical agreements from a theory offering a much less definite and specific picture of the interior of a ferromagnetic body than even Ewing's. Such agreements, nevertheless, emerge from the theory of Langevin and Weiss.

Langevin took as his point of departure the theory of temperature developed by the great savants Maxwell and Boltzmann (the same from which, by the way, the quantum-theory arose through the modifications made by Planck). To introduce as much, or as little, of this theory of temperature as is required for our present purpose, we envisage a sample of oxygen gas, N molecules per unit volume, in thermal equilibrium at absolute temperature T. Let each molecule be visualized as a rigid body of mass m, having three principal axes of rotation and corresponding moments of inertia I_1, I_2, I_3. The molecules are darting to and fro, with translatory velocities which may be specified by giving the three components u, v, w of each in some coordinate-frame. They are likewise revolving, with angular velocities which may be specified by giving the three components r, s, t of each along the principal axes of the molecule in question. The kinetic energy of the molecule is given by

$$\begin{aligned} K &= \tfrac{1}{2}mu^2 + \tfrac{1}{2}mv^2 + \tfrac{1}{2}mw^2 + \tfrac{1}{2}I_1r^2 + \tfrac{1}{2}I_2s^2 + \tfrac{1}{2}I_3t^2 \\ &= K_u + K_v + K_w + K_r + K_s + K_t, \end{aligned} \tag{1}$$

each of which six terms may be regarded as the kinetic energy associated with the variable which its subscript denotes. We will further suppose that each molecule is a magnet of moment M. When the gas is pervaded by a magnetic field H, each molecular magnet has a potential energy V_θ given in terms of the variable θ, the angle which its axis makes with the field, by the equation

$$V_\theta = -MH\cos\theta. \tag{2}$$

I propose now to show that Langevin's theory of magnetization is obtained by applying to the potential-energy term V_θ the same mode of reasoning as is customarily and familiarly applied to the kinetic-energy terms $K_u \cdots K_t$.

It is well known that the average kinetic energy of translation, the average of the sum of the terms K_u and K_v and K_w, taken over all the molecules of a gas of absolute temperature T, is proportional to T; it is, in fact, given by the equation

$$\overline{K_u + K_v + K_w} = \frac{3}{2}kT, \tag{3}$$

in which k stands for the ratio of the gas-constant R to the Loschmidt number N_0 (number of molecules per gramme-molecule).* The average of each of these three terms separately is equal to $\frac{1}{2}kT$; and this result was generalized by Maxwell and by Boltzmann to the three rotational terms in the expression for K, so that

$$\overline{K}_u = \overline{K}_v = \overline{K}_w = \overline{K}_r = \overline{K}_s = \overline{K}_t = \tfrac{1}{2}kT. \tag{4}$$

We go one step further in the analysis of the motion of the molecules. Consider the distribution-function for any one of these six variables, u for instance; it is given by Maxwell's formula:

$$dN = NC_u \exp\left(-\tfrac{1}{2}mu^2/kT\right)du, \tag{5}$$

in which dN stands for the number of molecules (among the N molecules occupying unit volume) for which the velocity-components along the x-axis lie between the values u and $u + du$. The constant C_u is so adjusted that the integral of dN over the entire range of values of u shall be equal to N; on being computed it turns out to be $\sqrt{m/2\pi kT}$. The quantity $\frac{1}{2}mu^2$ is the one hitherto designated as K_u. For the distribution-function with respect to u, which is the coefficient of du in (5), and may be denoted by $F(u)$, we therefore have:

$$F(u) = N\cdot\sqrt{m/2\pi kT}\cdot \exp.\ (-K_u/kT) \tag{6}$$

* The primitive way of deriving (3), reproduced in all elementary texts, is as follows: Imagine a cubical vessel one cm. along each edge containing N molecules; suppose that $N/3$ molecules are moving in lines parallel to each edge, with uniform speed v; each face is then struck with $Nv/6$ impacts per second, and in each impact an amount of momentum $2mv$ is communicated to the face, so that the average pressure upon the surface is $p = Nmv^2/3$. According to the well-known gas-law, $p = \rho RT/M$ (ρ standing for the density, M for the molecular weight of the gas); hence $Nmv^2/3 = \rho RT/M$, and recalling that $\rho = Nm$ and that $M/m = N_0$ and that $\frac{1}{2}mv^2$ is the kinetic energy K of a molecule, we have $K = 3RT/2N_0 = 3kT/2$. The same result is reached by more sophisticated methods of averaging.

and the distribution-functions with respect to v, w, r, s, and t differ only by the substitution of the appropriate kinetic-energy term for K_u, and (if necessary) of I_1 or I_2 or I_3 for m.

For the distribution-function with respect to θ, we shall write an equation copied after (5), as follows:

$$\begin{aligned} dN &= NC_\theta \, exp \, (- V_\theta/kT) \sin \theta d\theta \\ &= NC_\theta \, exp \, (MH \cos \theta/kT) \sin \theta d\theta. \end{aligned} \quad (7)$$

The constant C_θ is to be so adjusted that the integral of dN over the entire range of values of θ (which extends from 0 to π) shall be equal to N. It turns out that

$$C_\theta = a/(e^a - e^{-a}) = a/2 \sinh a \quad (8)$$

in terms of the parameter

$$a = MH/kT, \quad (9)$$

which we shall use often enough to justify the special symbol for it. The factor $\sin \theta$ in equation (7) requires comment. Imagine all the molecular magnets brought together at a point P, and their axes prolonged until these intersect a sphere of unit radius traced around P as center. The locus, upon this sphere, of the points of intersection of lines associated with magnets inclined at angles between θ and $d\theta$ to the field is a belt or collar of area $2\pi \sin \theta d\theta$. There are dN of these points, and they are distributed over this belt with surface-density $dN/2\pi \sin \theta d\theta$. By making dN proportional to the product of $\sin \theta$ into an exponential function, we make that surface-density, which is the density-in-solid-angle of the directions of the magnetic axes, proportional to the exponential function itself; and this is what is done.

We proceed to calculate the net magnetic moment of the assemblage of N molecular magnets. Resolving the moment of each, we find $M \cos \theta$ for its component parallel to the field-direction (with the perpendicular component we are not concerned, since the average of its values for all the molecules is obviously zero). Summing the values of these parallel components for all the molecules, we have:

$$I = \int_0^\pi M \cos \theta dN, \quad (10)$$

the symbol I being used for the sum of the parallel components, since this sum is precisely the intensity of magnetization per unit volume defined near the beginning of this article. Remembering (7) and (8), and performing the integration, we arrive at

$$\frac{I}{I_{\text{max.}}} = \coth a - \frac{1}{a} = L(a), \tag{11}$$

the symbol $I_{\text{max.}}$ being used for NM, the total magnetic moment which the assemblage of N molecules would have if they were all directed perfectly parallel to the field.

This function $L(a)$ is represented by the curve of Fig. 12, which departs from the origin with slope $a/3 = MH/3kT$, and bends over

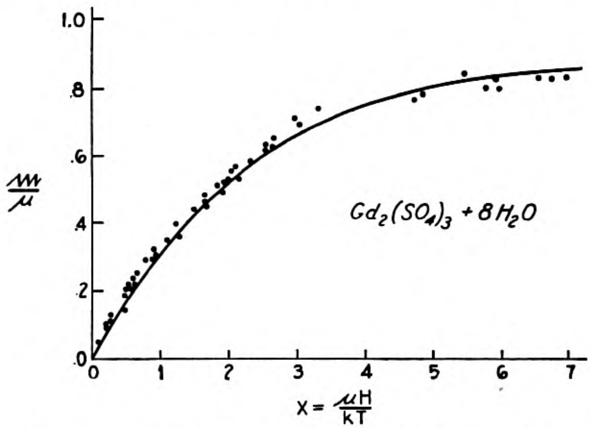


Fig. 12—The I-vs.-a curve of Langevin's theory of paramagnetism, and the data for gadolinium sulphate. (After P. Debye.)

toward its asymptote $L = 1$ without passing through any point of inflection. It has thus a resemblance to the initial curve; but one must not be misled by this, for Langevin's theory is not a theory of ferromagnetism. It is based on assumptions appropriate to a gas, and gases are not ferromagnetic; it gives no account of remanence, and remanence is an essential feature of ferromagnetic bodies. With a gas of which the molecules are permanent magnets, we should expect I to vary with H in the manner indicated by the curve.

Now as a matter of fact, in oxygen and other paramagnetic gases I is apparently proportional to H, up to the greatest fieldstrengths which can be applied:

$$I = \sigma_0 H. \tag{12}$$

This however does not necessarily mean that equation (11) is not valid; it may mean simply that the greatest available fields (some tens of thousands of gauss) are not great enough to pass beyond the sensibly-straight initial portion of the curve. If so, then

$$\sigma_0 = I_{\text{max.}} M/3kT = NM^2/3kT \tag{13}$$

and σ_0, the susceptibility of the material, should vary inversely as the absolute temperature. This, as Curie found, is true for the paramagnetic gases. It is true also for a number of salts in dilute solutions, and even for a certain number of solid substances, although for these the underlying assumptions would scarcely be expected to remain valid; one has the feeling that the data are left floating in the air by the withdrawal of the logical basis for the theory with which they agree.

Suppose nevertheless that the theory remains valid; then, for any substance of which the susceptibility σ_0 varies inversely as T, one can calculate the moment M of its molecular magnets from (13); for k is a known constant, and N is knowable at least when one is dealing with a gas of known density or a solution of known concentration (with solids there may be doubt as to the number of atoms grouped together to form an "elementary magnet"). Multitudes of such values have been computed; their orders-of-magnitude are 10^{-18} to 10^{-20}. Commonly they are expressed as multiples of a certain unit, the "Weiss magneton," which is equal to $1126/N_0$ or about $1.858 \cdot 10^{-21}$. Many of them are nearly integer multiples of this unit.*

On taking any observed value of M, and multiplying it by the corresponding value of N to obtain the "theoretical" value of $I_{max.}$ for the substance in question, we find that as a rule this last is so much larger than the highest value of I attained with practicable fields that there is no contradiction between the theory and the fact that I is sensibly proportional to H all through the feasible range of fieldstrengths. There is only one substance (gadolinium sulphate) for which $I_{max.}$ can be approached and this only at extremely low temperatures, below 5° absolute; in Fig. 12 the data are displayed; it is evident that the Langevin curve, drawn with the initial slope best suiting the points near the origin, fits fairly well to all the other points.

I pass now to the assumption whereby Weiss so extended Langevin's theory that it became competent to describe not only these simplest cases of paramagnetism in which $1/\sigma$ is proportional to T, but also the much more numerous cases of paramagnetic substances conforming to a more general law, and certain aspects of ferromagnetism also.

Formally the extension amounts to this, that in the expression for the parameter a which figures in equation (9), the fieldstrength H is replaced by a linear function of H and I:

* To enter into the long and fiercely debated questions about the meaning and even the reality of the Weiss magneton would lead me too far afield; but it is so frequently used as a unit in stating data of experiment that one must know at least its value.

$$a = M(H + nI)/kT, \tag{14}$$

which is transported bodily into the function $L(a)$ of equation (11). This is a very abstract way of putting the fact; but the more concrete ways have not been satisfying. One may say that the true field acting within the material is not H, but $(H + nI)$—that the actual though unverifiable field acting at any point in the inaccessible interior of the magnet is the sum of the field H_e due to objects in the external world, and the field H_i due to the "demagnetizing effect of the poles," *and an additional term proportional to the intensity of magnetization at the point in question.* The suggestion of Weiss, then, is tantamount to making a new assumption concerning this tantalizing internal field. The natural next step is, to visualize or explain the agent of the extra force, the "molecular field" as Weiss calls it; that is the step which no one has yet succeeded in making, not at least with general assent.

Making the expression in (14) the argument of $L(a)$, we see that the fundamental equation (11) now has the variable I in both its members, and must be solved for I. The resulting function is one of the infinitely many which have neither names nor well-known features, and most of those who write on this subject recommend the high-school expedient of plotting the curves representing the two functions

$$I = (kT/nM)a - H/n, \tag{15a}$$

$$I = I_{\text{max.}}L(a) = NML(a), \tag{15b}$$

in a coordinate-plane with I as ordinate and a as abscissa, and looking for the point or points of intersections between the two curves. These, which I shall designate for a few paragraphs as "the line" and "the

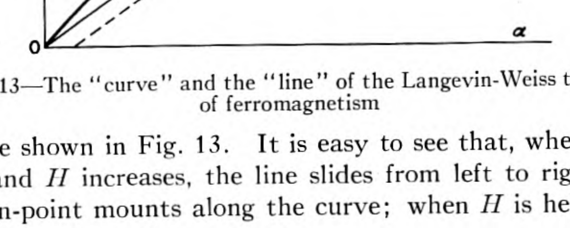


Fig. 13—The "curve" and the "line" of the Langevin-Weiss theory of ferromagnetism

curve," are shown in Fig. 13. It is easy to see that, when T is held constant and H increases, the line slides from left to right and the intersection-point mounts along the curve; when H is held constant and T increases, the line wheels counterclockwise around the point where it cuts the axis of abscissæ, and its intersection with the curve descends along the latter.

There is a valuable approximation, which is more nearly valid, the higher the temperature and the lower the field. At the origin, the tangent to the curve ascends with slope $NM/3$ (as I have said) and so long as a is not greater than unity, the ordinate of the curve agrees within six per cent with the ordinate of the tangent. If H is so small and T so great that the crossing of the line and the curve occurs within this range, the problem of locating it may be translated for all practical purposes into the algebraic problem of solving the simultaneous equations

$$I = (kT/nM)a - H/n, \qquad I = NMa/3, \tag{16}$$

achieving which, one obtains

$$I/H = \sigma = \frac{C}{T - \Theta}, \qquad \begin{cases} C = NM^2/3k \\ \Theta = nC. \end{cases} \tag{17}$$

The susceptibility of an assemblage of elementary magnets, in thermal equilibrium under the influence of an applied field on which there is superposed an extra field proportional to the magnetization of the assemblage, should then depend on temperature approximately

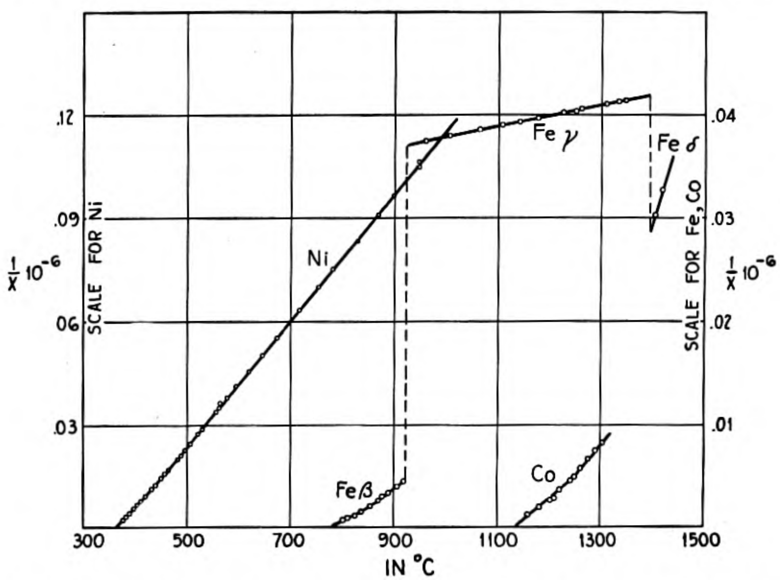


Fig. 14—Susceptibility-vs.-temperature curves for iron, cobalt, and nickel above their respective Curie-points. (After P. Debye.)

according to (17); the approximation being closer, the higher the temperature and the lower the field.*

Now there is a very large class of paramagnetic substances of which the susceptibilities at low fieldstrengths conform, over wide ranges of temperature, to equations like (17); and what renders the theory important for our present purposes is, that the ferromagnetic metals at high temperatures enter into that class. To make the test for any substance it is best to plot $1/\sigma$ as a function of absolute (or Centigrade) temperature. On doing this for nickel beyond the Curie-point (near 360° C.) one finds a curve which at first is somewhat bent, but beyond 410° passes into a beautiful straight line which continues undeflected to 900°. This line is shown in Fig. 14, together with data for iron beyond its Curie-point at 775°; among these, the points for temperatures between 920° and 1395° lie along a straight line which is sharply broken off at each end of that interval, being followed beyond 1395° by what seems to be the beginnings of an entirely different line, and preceded before 920° by a series of points which are well fitted by a pair of straight lines connected with each other at 828°. The data for cobalt beyond its Curie-point at 1130° likewise conform to a pair of connecting straight lines.

For each of these straight lines one may compute the values of the constants called C and Θ; and from these, if one accepts the theory, the values of the moment M of the elementary magnets and the coefficient n of the postulated extra force. In calculating M it is necessary to make an assumption about the number of elementary magnets per unit volume of the metal; assuming that there are as many such as there are atoms, and expressing M in Weiss magnetons, Weiss obtained the values 20.9, 17.4, 28.2 and 7.05 for the four straight lines of iron (in order of increasing temperature); 15.9 and 14.55 for those of cobalt; 8 for the solitary straight line of nickel. All these are of the orders of magnitude customarily found in dealing with paramagnetic gases and salts and solutions. The corresponding values of Θ are 1047, 1063, $-$ 1340, 1543; 1404, 1422; and 645. The corresponding values of n (which is the quotient of Θ by C) are of the order of several thousands. The postulated extra field must therefore be supposed enormously greater than the field H, and even the induction is quite insignificant by comparison with it. In one of these cases (and in many others among the paramagnetic salts,

* I should state that formulae of the same type as (17) may be derived without assuming that there is a molecular field, provided that we suppose that the distribution-in-energy of the atoms in thermal equilibrium is governed not by the equipartition-law, but by a quantum-law involving a zero point energy.

and in that of liquid oxygen) it must even be supposed *antiparallel* to the field H; for Θ is negative, and consequently so is n. Necessities such as these make it hard to accommodate the "molecular field" to what is known or conjectured about the interior of solid bodies.

Since it is necessary to assign several distinct values to the coefficient M in order to explain the behavior of iron over various ranges of temperature, one cannot maintain that the iron atom possesses a constant and characteristic magnetic moment which is the source of ferromagnetism. Any such notion, of course, would have been destroyed by facts already mentioned; but it is useful to know these in addition. Changes in M sometimes coincide with great and striking changes in the condition of the metal; at 920° iron exchanges its body-centered lattice (spacing 2.88A) for a face-centered lattice (spacing 3.60A) which it retains as the temperature rises until 1395° is attained, whereupon it returns to the body-centered-cubic arrangement. These alterations in atom-lattice are attended by changes in the physical properties of the metal, so great that three separate "modifications" of iron were distinguished and named before ever the atom-lattices were known or suspected: β-iron normally existing from the Curie-point to 920°, γ-iron from 920° to 1395°, δ-iron from 1395° upward. By certain processes these modifications may be enabled to survive in temperature-ranges not appropriate to them, but that is too long a story for these pages. Changes in M sometimes occur quite unaccompanied, so far as can be made out, by changes in atom-lattice or other physical features. The variation occurring at 828° in iron is of this type, and so is a mysterious change in nickel which in occasional samples brings about values of M near 9, instead of the usual 8 Weiss magnetons.

We turn to residual magnetism, on its explanation of which every theory of ferromagnetism must stand or fall. It is the supreme merit of the theory of Weiss that residual magnetism figures as a property which substances paramagnetic at high temperatures naturally and gradually acquire, when they are cooled below a certain critical point. We shall see this best by returning to Fig. 13. Begin by imagining the line corresponding to a particular pair of values of H and T; leave T constant, reduce H steadily to zero; the intersection of curve and line slides down the curve, *reaching the origin if the slope of the line is greater, stopping short of the origin if the slope of the line is less, than the slope of the tangent to the curve at the origin.*

The slope of the line is kT/nM; the slope of the tangent is $NM/3$; the critical condition is, that these be equal, and this occurs when

$$T = nNM^2/3k = \Theta,$$

i.e., when the temperature assumes the value of that constant Θ which previously entered into our equations. If T is greater than Θ, there should be no residual magnetism. If T is adjusted to be equal to Θ and then reduced gradually to zero absolute, the residual magnetization given from the theory—the ordinate of the point where the curve is intersected by the line of slope kT/nM passing through the origin—increases continuously from zero to its limiting value NM, following the curve traced in Fig. 9. That is the central idea of Weiss' theory of ferromagnetism.

The first of the predictions from the theory which can be put to test is the equality between the temperature at which residual magnetism disappears—the Curie-point—and the constant Θ in the equation (17) for the paramagnetism of the substance beyond the Curie-point. For nickel, the agreement is good: 633° against 645° absolute. For cobalt and for iron, the first short straight line out of the sets of two and four respectively, which are given for these metals in Fig. 14, is so adjusted that Θ agrees perfectly with the Curie-point; its aptness to the plotted data supports the theory.

The next question to be asked is whether the curve of Fig. 9 corresponds to experience. In analyzing this question, one makes the discomfiting discovery that the quantity which was defined as residual magnetization in the theory cannot be identified with the quantity defined as remanence in describing the experimental hysteresis-loops. This results from an imperfection, or at least an incompleteness, in the theory. There is nothing in it to account for the initial curve; there is nothing to account for the gradual increase in I produced by applying a gradually-increasing field to an initially-demagnetized piece of iron, and in fact there is nothing to account for the existence of demagnetized pieces of iron at all—every block of iron at a temperature below Θ should possess, whenever it is not under the influence of an external field, the residual magnetization calculated from the intersection-point of the curve $NML(a)$ and the line of slope kT/nM which passes through the origin.

On grasping this situation, one is likely to feel that the theory has collapsed. The situation can be saved, however, by supposing that the "demagnetized" metal subdivides itself into a vast number of little regions, zones, or filaments, each of which possesses the full residual magnetism of the theory, while in direction their magnetic moments are oriented quite at random. It is not possible to identify these with individual crystals, nor with any other discernible granulations of the metal. Perhaps they are to be identified with the chains of elementary magnets once postulated by Ewing; it would be gratifying to make a connection between the theories of Ewing and Weiss. Perhaps they are the units from which arise the separate clicks which constitute the Barkhausen effect. As for the initial curve, attempts must be made to explain it either by supposing that the increasing field wheels the magnetic moments of the several zones gradually into parallelism with itself, or—what is more probable—that the field abruptly reverses, one after the other, all the magnetic moments which initially are inclined to it at angles superior to 90°. By suitably combining these two images, one may copy almost any possible form of initial curve. I cannot enter into these questions, except to answer as far as possible what I designated as the second question to be asked in testing the theory: what observable quantity is to be compared with the "residual magnetization" predicted from the theory of Weiss?

A piece of iron brought to saturation by a large applied field is supposed to consist of these magnetized zones, their moments all directed either parallel or at least at inclinations of less than 90° to the field. The applied fieldstrength should elevate the magnetization of each to a value somewhat greater (corresponding to an intersection-point somewhat farther along the "curve" of Fig. 13) than the predicted "residual magnetization"; but the values of n and I and hence their product are so enormous that the addition is only slight. The saturation intensity of magnetization of the iron, $I_{max.}$, should then be very nearly equal to the predicted residual magnetism, if all the magnetic moments are parallel; or to one half of the predicted residual magnetization, if the magnetic moments are distributed at random over the directions inclined at less than 90° to the applied field. In the former case, the variation of $I_{max.}$ with T should follow the curve of Fig. 9; in the latter case, a curve of the same form. The actual observations upon iron, nickel, cobalt and magnetite are shown in that figure, and the reader may judge of the agreement for himself.

Comparison of Ewing's Theory with that of Langevin and Weiss

At first glance the Ewing model and the Langevin-Weiss conception of a ferromagnetic substance seem extremely different; contradictory, in fact. In Ewing's view, the perpetual effort of the applied field to align the elementary magnets is hindered by the forces which these exert on one another. In Langevin's theory, the antaonigst of the applied field is the thermal agitation. Now Langevin's theory is competent to deal with paramagnetic substances which are difficult to magnetize, but not with iron and the like which are strongly magnetized by weak fields. This means that the thermal agitation is too strong an antagonist to the applied field. Weiss therefore provided the latter with a powerful ally, in the form of an intense molecular field parallel to it and proportional to the magnetization. The applied field and its ally together are able to overpower the thermal agitation and bring about saturation in cold iron. Now to say "molecular field" is merely to use a different phrase for "influence of the atoms on one another." In the theory of Weiss, this influence of the atoms on one another helps the field to align them; in Ewing's theory, it hinders the field. How do away with this arrant contradiction?

Perhaps a partial union may be effected, in this wise. According to Langevin and Weiss, a piece of cold iron consists of a multitude of small zones or regions of atom-groups, each magnetized to a high degree, their directions of magnetization dispersed at random; an applied field acts primarily by wheeling these magnetizations into line. According to Ewing, a piece of cold iron consists of a multitude of chains or pairs of systems of elementary magnets, which an applied field upsets, perhaps only to re-weld them anew into more favourably oriented chains. Weiss deals with the state of affairs inside the atom-groups; Ewing deals with the effect of the applied field in breaking up and rebuilding the atom-groups. Might one say that Weiss explains the conditions, under which the elementary magnets form themselves into groups or chains such as Ewing preassumed? that Ewing describes the action of the external field upon these groups, an action which Weiss left imprecise? so that the two theories, when properly revised, will complement each other? It seems possible. At all events, each of the theories has so many successes to its credit, that there can be no thought of discarding either for the sake of the other. Those who are weary of trying to reconcile waves and quanta might refresh themselves by reflecting on this problem.

McKeehan's Theory

In the theory of McKeehan, magnetostriction is promoted to the dominant role. The distortion which a metal undergoes when it is magnetized is held responsible for hysteresis, and for the fact that the rise of the I-vs.-H curve is gradual, not sudden. This view was suggested by the fact which I have mentioned already: that, in the series of the permalloys, the permeability reaches a surprisingly high maximum value and the hysteresis a surprisingly low minimum value, just at that alloy of which the magnetostriction is indetectably small until saturation is nearly attained—the alloy intermediate between those which lengthen and those which shorten when magnetization commences. The alloy which is most rapidly magnetized when the field is gradually increased from zero, and which dissipates the smallest amount of energy when the field is varied in cyclic fashion, is also precisely the one which suffers the least deformation. From this McKeehan drew the inference, that were it not for the deformation inseparable from the act of magnetizing, the initial curve for every metal would rise swiftly from the origin to saturation, and the sides of the hysteresis-loop would fall together.

D. The Atomic Magnets

Had I announced at the beginning of this article that some sixty pages would be spent over the data of ferromagnetism and the theories of the influence of elementary or atomic magnets on one another, and only a few closing paragraphs over the atoms which are supposedly responsible for the whole affair, the plan might have seemed most ill-adjusted to the relative interest of these divisions. Now, I hope, it will seem less perverse. The truth is, that we do not understand ferromagnetism well enough to draw from it any reliable conclusions concerning the atomic magnets. For these, we must consult the behavior of paramagnetic substances, and line-spectra, and the observations of Gerlach and Stern and their followers upon streams of atoms flying through magnetic fields.

In the apparatus of Gerlach and Stern, the atoms are probably as nearly free from mutual forces as atoms in the laboratory can ever be; having issued from a small hole in the wall of a furnace full of hot vapor, they rush swiftly across a high vacuum while they are being examined. In the mapping of absorption-spectra, the atoms are those of a rarefied gas, and are "free" in the sense in which atoms of gases are free—that is to say, they are influenced only by those agencies which establish and maintain thermal equilibrium, agencies which we commonly conceive as short, sharp collisions between atom and atom. Some paramagnetic gases behave toward an applied magnetic field as though their molecules, some salt-solutions behave as though their ions, were magnets of fixed permanent moment on which the field can act, but otherwise were free in the foregoing sense. Other gases and salt-solutions behave as though their molecules or ions were permanent magnets, influenced by the applied magnetic field and by an extra field proportional to the magnetization of the assemblage, and otherwise free except for the agencies which establish thermal equilibrium and maintain it.

In all the foregoing cases of atoms or molecules or ions enjoying variously close approximations to perfect freedom, the theories are good enough to make it possible to bring about quantitative agreement between theory and experiment, simply by choosing appropriate values for the magnetic moments of these particles. The values so determined nearly always lie between 10^{-18} and 10^{-20} C.G.S. units.

Ferromagnetic substances are solids, and we need not be surprised that the mutual influence of the atoms becomes so great as to make the task of devising a theory much more difficult. Ewing, it is true, did show that elementary magnets of a particular shape and crowded close together would form systems displaying the peculiar features (hysteresis, and a crooked magnetization-curve) of ferromagnetics. Weiss did show that atomic magnets, subject to the agencies which bring about thermal equilibrium and maintain it, and in addition to a field proportional to the magnetization of the assemblage and enormously great, would form systems displaying residual magnetism below a certain temperature, and paramagnetic above. Dazzling as these achievements are, the theories are not so good that they can be brought into complete accord with the data, simply by choosing appropriate values for the moments of the imagined elementary magnets.

Can we at least assign a value of the order familiar among paramagnetics, 10^{-19} for instance, to the magnetic moment of (say) the iron atom—that is to say, the atoms of a piece of solid pure iron, since iron is not in all conditions ferromagnetic—without definitely contradicting any fact of experience? Probably we can. In fact, the saturation-values of the magnetizations of iron, nickel, and cobalt support this idea. If saturation signifies that all the atomic magnets are parallel, then the magnetic moment of each must be the quotient of $I_{max.}$ by the number of atoms in unit volume; at all events, the magnetic moment of the atom cannot be less than the quotient, by that number of atoms, of the highest value of I ever observed. Now the highest values of I are observed at the lowest temperatures; extrapolating from the data (shown in Figure 9) to zero absolute, Weiss obtained values of the quotient which are indeed of the order 10^{-19}—eleven "magnetons" for iron and three for nickel, and probably eight for cobalt. This concordance with the values of magnetic moment to which we are accustomed among free atoms is evidently important. However, as Ewing found, we cannot take the natural next step of supposing that each atom is a long slender magnet having its ends very close to the ends of the adjacent magnets; for then the I-vs.-H curve of the assemblage would not agree with the initial curves observed in practice.

Everyone now agrees with the idea, proposed more than a century ago by Ampère, that atoms are magnets because of the circulating charges which they contain. The estimates of atomic moments deduced from line-spectra are based on this assumption, and the verified correctness of these estimates sustains it. Now, if a magnetic atom is a whirl of electricity, it possesses angular momentum as well as magnetic moment. If so, the process of magnetizing an iron wire involves the bringing-into-parallelism of myriads of spinning-tops, of which the angular momenta when all aligned combine into a respectable sum. If this goes on inside a wire during magnetization, there should be a "recoil" somewhere, comparable to the recoil of a gun when a shell is fired—the suspension of the wire should receive an opposite angular momentum, experience a torque. Conversely, the process of twisting an unmagnetized wire should impress a lateral torque upon myriads of spinning-tops of which the axes point in directions scattered at random; each of these should be urged to set itself more nearly parallel to the axis of the twist, which is the axis of the wire; and the twisting should therefore magnetize the wire.

Both of these effects, which jointly are called the "gyromagnetic effect," have been detected and measured. From the measurements (thus far performed upon iron, nickel, cobalt, magnetite and a Heusler alloy), it results that the ratio of the angular momentum P to the magnetic moment M of an elementary magnet conforms to the equation:

$$P/M = mc/e,$$

in which m stands for the mass of the electron and e/c for its charge measured in electromagnetic units. *This is the value which would be expected for the ratio, if the elementary magnet is an electron spinning upon itself.*

Now there are weighty reasons for supposing that the conception of a "spinning electron," possessing a fixed characteristic angular momentum and a permanent magnetic moment e/mc times as great, may be what is required to complete the theory of line-spectra of free atoms which Bohr began. The gyromagnetic effect of the ferromagnetic solids therefore indicates that the elementary magnets scattered through these are the same as the elementary magnets located in free atoms—they are electrons, or groups of electrons suitably linked together. The test cannot be made upon paramagnetics, for they cannot be (or at least have not yet been) strongly enough magnetized. Ferromagnetic substances are the only ones which in a feasible field acquire so great a magnetization that the recoil from the spinning electrons is detectable. This seems to be as yet the only contribution of ferromagnetism to contemporary atomic theory.

Yet even if we take it for settled that the elementary magnets within the atoms of a solid piece of iron are spinning electrons, the real problem of ferromagnetism remains unsolved. If the elementary magnets in iron are just like those in all other atoms, how does it happen that iron and two other elements alone may be ferromagnetic? that even iron may cease to be ferromagnetic, if mixed with a little manganese? that manganese and copper and aluminium can become ferromagnetic when and only when alloyed together? Since apparently we must not suppose that each atom of iron is distinguished from all those of never-ferromagnetic substances through having a peculiar kind of magnet inside it, we must suppose that something strange in the arrangement of the electron-magnets of the iron atom permits it to be so distorted, and so to distort its neighbors, that on occasion its neighbors and itself jointly develop ferromagnetism. There is something extraordinary about the systems of 26 and 27 and 28 electrons about a nucleus, which iron and nickel and cobalt atoms are. Their individual electrons are not unique; by themselves, or as ions in a solution, they show nothing unique; but they turn into something unique when they are rightly compounded together into a solid. The theories of ferromagnetism and the gyromagnetic effect have limited without solving the fundamental problem of ferromagnetism: what is it that makes the difference between the ferromagnetic substances, and all the rest?

Acknowledgments and References

The foregoing article is based largely upon the books of J. A. Ewing (*Magnetic Induction in Iron and Other Metals;* Electrician, 1900), P. Weiss and E. Foex (*Le Magnétisme;* Colin, 1926) and E. C. Stoner (*Magnetism and Atomic Structure;* Methuen, 1926); the articles by S. Bidwell in the eleventh edition of the Encyclopædia Britannica, by P. Debye in volume 6 of the *Handbuch der Radiologie*, by E. Gumlich and R. Gans in *Die Kultur der Gegenwart*, by K. Honda in the *Dictionary of Applied Physics;* and the articles of L. W. McKeehan on ferromagnetism (*Journ. Franklin Inst.* **197**, pp. 583–602, 757–786; 1924), magnetostriction (*ibid.* **202**, pp. 737–773; 1926) and the permalloys (*Phys. Rev.* (2) **28**, pp. 146–166; 1926, and others there cited).

Some of the very recent papers upon magnetization of single crystals are those of W. L. Webster (*Proc. Roy. Soc.* **A107**, pp. 496–509; 1925); K. Honda and S. Kaya (*Tohoku Univ. Sci. Rep.* **15**, pp. 721–753; 1926); W. Gerlach (*Z.S. f. Phys.* **38**, pp. 828–840; 1926). For magnetostriction of single crystals, see W. L. Webster (*Proc. Roy. Soc.* **A109**, pp. 570–584; 1925) and K. Honda and Y. Mashiyama (*Tohoku Univ. Sci. Rep.* **15**, pp. 755–776; 1926). For the data concerning permalloys see,

in addition to the papers already cited, that of H. D. Arnold and G. W. Elmen (*Journ. Franklin Inst.* **195**, pp. 621–632; 1923). The gyromagnetic effect, the data and the theories of paramagnetic substances, and diamagnetism are treated very fully in the above-cited book of Stoner; paramagnetism, and the interesting and important magneto-caloric effects which I had not space to discuss, in the book of Weiss and Foex; diamagnetism in a late article by E. S. Bieler (*Journ. Franklin Inst.* **203**, pp. 211–242; 1927).

I am much indebted for the comments and counsel and information abundantly given by Dr. L. W. McKeehan during the preparation of this article, and for the opportunity of using several cuts prepared for his article "Ferromagnetism"; to Dr. O. E. Buckley for reading and commenting upon a great part of the manuscript; and to Mr. L. A. MacColl for much collaboration in studying the equations of Ewing's model.

Abstracts of Bell System Technical Papers Not Appearing in this Journal

The Crystal Structure of Magnesium Platinocyanide Heptahydrate.[1] RICHARD M. BOZORTH and F. E. HAWORTH. Positions of the Mg and Pt atoms in crystals of $MgPt(CN)_4 7H_2O$. These have been definitely determined by means of x-ray oscillating-crystal photographs and Laue photographs, using the theory of space-groups. Because the other atoms are too light in comparison with the metal atoms, especialy Pt, their positions could not be determined. The Pt atoms are located at 0 0 0 and $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$, the Mg atoms at 0 0 $\frac{1}{2}$ and $\frac{1}{2}$ $\frac{1}{2}$ 0, in a tetragonal unit of structure 14.6A $\times$ 14.6A $\times$ 3.13A. Two units of structure are shown in the figure. The peculiar optical properties are believed to be associated with the unusual arrangement of the heavier atoms in widely spaced rows parallel to the tetragonal axis. In these rows Mg atoms alternate with Pt atoms, and the distance between any two adjacent atom-centers is 1.57A. The shortest distance between rows, however, is 10.3A, 6.6 times the distance between atoms in the same row. The atomic radii of Mg and Pt as determined by Bragg from other crystal data do not agree with the observed distance between these atoms, the calculated value being 2.7A, the observed distance 1.57A. The observed distance, however, is consistent with that calculated by the method of Davey, who assumes that the radius of an ionized atom differs much from the radius of the same atom un-ionized, and that the radii of Cs^+ and I^- are substantially equal in crystals of CsI.

Photoelectric Emission as a Function of Composition in Sodium-Potassium Alloys.[2] HERBERT E. IVES and G. R. STILWELL. The entire series of alloys of sodium and potassium have been investigated with respect to the relative values of the photoelectric currents produced by light polarized with the electric vector in and at right angles to the plane of incidence. The pure metals when molten exhibit values below three for the ratio of the two emissions; the alloys show three maxima at compositions approximately 20, 50, and 90 atomic per cent of sodium, with values from 10 to 30 for the ratio; the minima between show low values approximating those for the pure metals. The maxima and minima of the ratio of emissions are due to complicated variations in magnitude of the two emissions compared.

[1] *Physical Review*, Feb. 1927, Vol. 29, No. 2, p. 223.

[2] *The Physical Review*, Feb. 1927, Vol. 29, No. 2, p. 252.

Submarine Insulation with Special Reference to the Use of Rubber.[3] R. R. Williams and A. R. Kemp. (1) Soft vulcanized rubber, though not well adapted to some of the processes of manufacture of submarine cable, can be so made as to be mechanically and electrically suitable and to withstand the action of sea water in a manner comparable with that of gutta percha over a period of a few years. Whether such rubber will retain these characteristics for decades remains to be demonstrated, but it seems probable that it will.

(2) The principal factor to be controlled in producing this result is the amount of water absorbed by the rubber.

(3) Osmotic pressure of internal and external fluids is of prime importance in governing the in-flow of water into rubber and gutta percha.

(4) Lowered water absorption is achieved by removal of water-soluble matter from the rubber, the choice of an insoluble, non-reactive filler of suitable particle size and having a minimum of adsorbed gases or other contamination on its surfaces.

(5) The electrical characteristics of rubber compounds and of gutta percha are clearly related to their water content but are not simple functions of the water content.

(6) It appears that the mode of distribution of water is also extremely important.

(7) Most fillers for rubber compounds are not suitable for submarine insulation, either because of undesirable intrinsic electrical properties or because they are conducive to changes incident to water absorption. Hard rubber dust, silica and zinc oxide are the best fillers from these standpoints so far as known.

An Efficient Apparatus for Measuring the Diffusion of Gases and Vapors through Membranes.[4] Earle E. Schumacher and Lawrence Ferguson. An efficient diffusion measuring apparatus, embodying a mechanical clamp and a mercury seal, is described. This apparatus can be used for measuring the rate of diffusion of gases and vapors through materials such as rubber, waxes, leathers and certain types of paper.

Investigation of the Thermionic Properties of the Rare Earth Elements.[5] Earle E. Schumacher and James E. Harris. Thermionic emission measurements over a range of temperatures were made on samples of pure Ce, La, Pr, Nd, Sa and the aluminum alloys of Yt, Eu, Ga,

[3] *Jr. of the Franklin Inst.*, Jan. 1927, Vol. 203, pp. 35–61.

[4] *J. Amer. Chem. Soc.*, Vol. 49, 427 (1927).

[5] *J. Amer. Chem. Soc.*, Vol. 48, pp. 3108–3117 (1926).

Tb, Dy, Ho, Er, Th, Yb and Lu. These measurements showed the rare earth elements, without exception, to be more active thermionically than the commonly occurring metals. At 1800° C. all of these metals gave emissions of more than 10^5 that of clean tungsten at the same temperature.

The Solidus Line in the Lead Antimony System.[6] EARLE E. SCHUMACHER and FOSTER C. NIX. An investigation of the solidus line above the solid solution field for the lead antimony system was made by the quenching test procedure. Three points were determined between the melting point of pure lead and the end of the eutectic horizontal. The position of the solidus line has been precisely fixed.

Production Control.[7] C. G. STOLL. This paper treats the subject of production control from the practical rather than the theoretical point of view. It is confined largely to a description of the generally accepted principles of production control as applied in the Manufacturing Department of the Western Electric Company. This plant employs approximately 30,000 people and produces annually over $150,000,000 worth of manufactured products. These products are comprised of some 13,000 kinds of apparatus containing over 110,000 different parts.

The paper discusses the organization of the factory, which is set up along functional lines, and also the extensive system of records and charts used to facilitate the work of the organization and to assist in production control.

The Significance of the Dielectric Constant of a Mixture.[8] HOMER H. LOWRY. It is pointed out that in many cases it would be of great value to be able to calculate either the dielectric constant of a mixture of substances of known dielectric constants or, knowing the dielectric constants of a mixture of two components and that of one of the components, to calculate the dielectric constant of the other. A review of the literature, however, shows that this can be rarely accomplished. This is due mainly to the inadequacy of the theories of dielectrics, all of which are insufficiently developed to include the dielectric behavior of mixtures. Nevertheless, as is shown, many attempts have been made to develop formulæ of theoretical significance for application to mixtures. Inspection of the derivation of these formulæ shows that those with the best theoretical background are limited to such special cases that they are of practically no value.

[6] A. I. M. E. Pamphlet No. 1636-E, Feb. 19, 1927.

[7] *Mechanical Engineering*, Vol. 49, p. 201, 1927.

[8] *Jr. of the Franklin Institute*, 203, 413–439, 1927.

A brief review of these formulæ is given together with a brief account of the results of experimental investigations on the dielectric behavior of mixtures. A rather extended bibliography is given.

The Effect of Moisture on the Electrical Properties of Insulating Waxes, Resins and Bitumens.[9] J. A. LEE and HOMER H. LOWRY. The results of measurements of dielectric constant and effective conductivity at 1,000 cycles and resistivity are reported for 31 waxes, resins and bitumens, including not only naturally occurring products but also commercial dielectrics and mixtures. The measurements were made on the materials initially in a thoroughly dry condition, after six months' immersion in a salt solution corresponding qualitatively to exposure to 98 per cent relative humidity, and after having been redried. All the insulating materials studied absorbed water under the conditions of experiment. The absorption was least with the hydrocarbons and greatest with shellac and bayberry wax. In general, the greatest increase in capacity and conductivity and the greatest decrease in resistivity were shown by the materials which absorbed the most water. The percentage change was much greater in the conductivity and resistivity than in the dielectric constant, as was to be expected.

The Mechanism of the Absorption of Water by Rubber.[10] H. H. LOWRY and G. T. KOHMAN. Data are reported which show the influence of the various factors which determine the amount of water absorbed by any given sample of rubber. From a consideration of the results obtained, it was concluded that, at a given temperature, the most important external factor determining the amount of water absorbed by a given sample of rubber is the vapor pressure of water with which it is in equilibrium. The data show further that the water-soluble constituents within the rubber are responsible for most of the water absorbed at high humidities, that increasing the rigidity of a rubber compound decreases greatly the amount of water absorbed, and that aging increases the water absorption. It is pointed out that all the experimental facts are consistent with the view that the absorption of water by rubber consists of two processes: the formation of a true solution of water in rubber and the formation of solutions internal to the rubber of the water-soluble constituents of the rubber which can be removed by washing.

[9] *Jr. of Industrial and Engineering Chem., 19*, 302–306, 1927.
[10] *Jr. of Physical Chemistry, 31*, 23–57, 1927.

Rapid Evaluation of Baked Japan Finishes.[11] E. M. HONAN and R. E. WATERMAN. The service life of a japan film baked on metal can be evaluated by determining the rate of decomposition of the film when it is placed in an 8.5 per cent phenol-water solution. The effect of the time and temperature of baking the film and the cleanness of the metal previous to applying the japan can also be evaluated. The 8.5 per cent phenol solution is a desirable testing solution because its composition is quite constant at ordinary room temperatures and is not changed by the evaporation of the water.

Magnetostriction. L. W. MCKEEHAN.[1] This paper contains the principal part of three lectures given at the Franklin Institute in April 1926. The history of investigations on the changes in dimensions which accompany magnetization and the changes in magnetization which accompany forcible changes in dimensions is sketched and a classification of the rather complicated cases which have been examined is offered. The bearing of magnetostriction on theories of ferromagnetism is emphasized and a number of new experimental results are described. A representative bibliography is appended.

[11] *Ind. and Eng. Chem.*, Oct. 1926, Vol. 18, pp. 1066–1068.
[1] *Journal of the Franklin Institute* **202**, 737–775 (1926).

Contributors to this Issue

J. R. Shea, B.S. in E.E., University of Wisconsin, 1909; Manufacturing Department, Western Electric Company, 1909–. Mr. Shea is now Assistant Superintendent of the Manufacturing Development Branch covering metal manufacturing and metallurgical work.

S. McMullan, M.A., McMaster University, Toronto, 1912, LL.B., Chicago Kent College of Law, 1924; Manufacturing Department, Western Electric Company, 1916–. Mr. McMullan is now Development Engineer in the Manufacturing Development Branch on the manufacture of copper rod and wire.

C. R. Moore, B.S. in Mechanical and Electrical Engineering, Purdue, 1907; E.E., Purdue, 1910; Instructor and Assistant Professor Electrical Engineering, Purdue, 1907–13; Manager of La Fayette Electric and Mfg. Co., 1913–14; Associate in Electrical Engineering, University of Illinois, 1914–16; Engineering Department of the Western Electric Co., 1916–24; Bell Telephone Laboratories, Inc., 1925–. Mr. Moore for several years has been associated with transmitter development work and has contributed important inventions relating to telephone instruments and acoustic devices.

A. S. Curtis, Ph.B., 1913, E.E., 1919, Sheffield Scientific School; Instructor in Electrical Engineering, Yale University, 1913–17; Engineering Department, Western Electric Company, 1917–24; Bell Telephone Laboratories, Inc., 1925–. Mr. Curtis' work has been connected with the development of telephone instruments.

A. G. Landeen, E.E., University of Minnesota, 1910; incandescent lamp development and manufacturing, General Electric Company, 1910–19; Engineering Department, Western Electric Company, 1919–24; Bell Telephone Laboratories, 1925–.

Ralph Bown, M.E., 1913, M.M.E., 1915, PhD., 1917, Cornell University; Captain Signal Corps, U. S. Army, 1917–19; Department of Development and Research, American Telephone and Telegraph Company, 1919–. Mr. Bown has been in charge of work relating to radio transmission development problems and recently has given particular attention to the quantitative and engineering side of transatlantic telephony.

W. P. Mason, B.S., University of Kansas, 1921; M.A., Columbia, 1924; Engineering Department, Western Electric Company, 1921–24; Bell Telephone Laboratories, Inc., 1925–. Mr. Mason has been associated with work on carrier current communication, and more recently with the development of filters and subscribers' sets.

Karl K. Darrow, S.B., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D. in physics and mathematics, University of Chicago, 1917; Engineering Department, Western Electric Company, 1917–24; Bell Telephone Laboratories, Inc., 1925–. Mr. Darrow has been engaged largely in preparing studies and analyses of published research in various fields of physics.