# The Bell System Technical Journal

*January, 1935*


## Wide Band Transmission Over Balanced Circuits *


By A. B. CLARK



In a recent paper [1] amplifiers capable of handling frequency band widths of the order of 1,000 kc. or more are described, together with terminal apparatus for effectively utilizing these wide bands for telephone, telegraph and television purposes. The paper confines itself to the coaxial line structure for transmitting these wide frequency ranges but points out that broad-band transmission is also applicable to balanced conductor systems. The present paper discusses briefly some of the possibilities of the more familiar balanced circuits, circuits more or less as they now exist in the present plant being first considered, following which are circuits obtained by new construction. Wide-band transmission over balanced circuits offers interesting possibilities both for circuits in the present plant and for new construction.


A HIGH degree of *electrical balance* [2] has been for a long time a fundamental requirement of telephone transmission lines. This has been required not only to prevent interference entering into telephone circuits from other types of electrical circuits but also to prevent mutual interference between the closely adjacent telephone circuits on open-wire lines or in cables.

In the central offices, to be sure, unbalance in apparatus has been frequently employed for simplicity and convenience. In toll circuits, however, such office unbalance has been electrically separated from the outside plant by the use of repeating coils or otherwise. In local circuits the high standard of balance required for toll circuits has not generally been necessary since the exposures to interfering fields are less severe and the range of speech levels is much smaller. However, when local circuits are connected to toll circuits the unbalances are kept electrically separated from the toll circuits by the use of repeating coils. In recent years the tendency to the use of higher frequencies in com-


* Published in January, 1935 issue of *Electrical Engineering*. Scheduled for presentation at Winter Convention of A. I. E. E., New York City, January 22–25, 1935.


[1] "Systems for Wide-Band Transmission Over Coaxial Lines" by L. Espenschied and M. E. Strieby, *Elec. Engg.*, October, 1934; *Bell Sys. Tech. Jour.*, October, 1934.

[2] As used here, the term *electrical balance* refers to the two sides of a telephone circuit. To secure such balance, the aim has been to construct the go and return conductors of each circuit of the same gauge and material and to locate them symmetrically with respect to earth and to surrounding conductors. The aim has also been to apply to each circuit terminal apparatus symmetrical with respect to its series impedances and shunt admittances to ground.

munication circuits, the increase in the strength of interfering fields and the development of highly efficient amplifiers has led to constantly more exacting requirements in electrical balance of telephone circuits.

The development of multichannel systems by carrier methods employing constantly increasing frequency ranges has placed particularly exacting requirements on such electrical balance. In a recent paper, "Carrier in Cable," [3] are described the balancing methods which have been developed to permit the use of an increased frequency range in such cables. The recently published paper, "Systems for Wide Band Transmission Over Coaxial Lines" [1] points out the possibilities and possible requirements for very much wider frequency ranges. In that paper, coaxial lines are proposed which are particularly interesting in that they abandon electrical balance altogether and depend entirely on metallic shielding.

For such wide frequency range transmission, very interesting and important questions are raised, first as to the extent to which such wide bands can be placed on existing types of structure which are based on balance and, second, as to whether new construction designed particularly for such wide bands should depend on balance or shielding alone or a combination of the two. It is the purpose of this present paper to discuss these questions.

As noted in the Espenschied-Strieby paper, the apparatus described for broad-band transmission on concentric structures would also serve for other types of line structures. There would, of course, be problems in either balancing the apparatus or isolating its unbalances from the line structure. For the purposes of the present paper it is assumed that there will be no important reaction from the apparatus standpoint on this consideration of line balance and shielding.

## Existing Cables

The attenuation of pairs in existing cables has a characteristic with respect to frequency generally similar to that of coaxial conductors but is, naturally, considerably higher because of the smaller physical dimensions and higher dielectric losses of the cable pair. For example, at a million cycles an ordinary 19-guage cable pair has an attenuation of about 18 db per mile, an ordinary 16-gauge pair about 14 db per mile, while a small-sized coaxial structure has an attenuation of about 6 db per mile. This means more repeaters for the cable pairs, three times as many for 19-gauge and a little more than twice as many for 16-gauge. Also, it means more difficulty in maintaining stability of transmission,

[3] A. B. Clark and B. W. Kendall, *Elec. Engg.*, July, 1933; *Bell Sys. Tech. Jour.*, July, 1933.

including overcoming of the variations due to the effect of temperature changes.

Stable and highly linear repeater gain can be produced so readily now-a-days, thanks to the negative feedback amplifier invented by Mr. H. S. Black,[4] that the idea of such high attenuations is no longer appalling even though on 16-gauge pairs it means repeaters spaced only about four miles apart. Overcoming the transmission variations due to temperature with automatic regulators introduces no fundamentally new problems, but, of course, the complexity and precision of regulation must be considerably higher due to the considerably larger variations.

Crosstalk, of course, must be given special consideration. First of all, it is necessary to restrict transmission of a given high-frequency band to only one direction in a single cable; the other direction must be supplied by another cable [3] or other separate transmission medium.

Considering transmission in one direction only, if only one pair in a cable is set aside for high-frequency transmission of, say, a million-cycle band, most, but not all, of the crosstalk difficulty can be avoided. However, the fact must be reckoned with that if one pair in a cable is singled out and an amplifier is applied having 60 db or more gain at a point intermediate between voice-frequency repeater stations, the amplifier will have a strong tendency to sing due to crosstalk between the pairs connected to the input and output and the other pairs in the cable. If two cables are available, this difficulty can be avoided by jumping from one cable to another every time the high-frequency amplification is introduced. If two cables are not available, overcoming the difficulty may call for the insertion of high-frequency choking devices in some or all of the low-frequency cable conductors at points where high-frequency amplification is introduced.

Considering now crosstalk between two circuits in a cable transmitting in the same direction, assuming the amplifier difficulty to have been overcome, tests have been made on various cables in the field from which these conclusions have been drawn. For telephone message purposes it is probably uneconomical to apply million-cycle frequency ranges to more than a single pair in an existing cable. However, with television, the crosstalk requirements are much less exacting. This is because the range of intensities necessary for a good television image is much less than is needed to accommodate message telephone subscribers and, therefore, a considerably larger ratio of extraneous current to maximum signal current can be tolerated. Tests indicate that two or more television channels, each a million

[4] "Stabilized Feedback Amplifiers," H. S. Black, *Elec. Engg.*, January, 1934; *Bell Sys. Tech. Jour.*, January, 1934.

cycles wide (possibly wider), can be transmitted over separate properly arranged pairs in the same direction in a single existing cable without serious disturbance due to crosstalk.

With respect to noise in existing cables, the matter of principal concern is noise produced in telephone offices by apparatus working on other circuits, since the natural shielding afforded by the cable sheath largely eliminates noise from outside sources. Two methods are available for control of the noise produced in telephone offices: (1) Introduction of high-frequency choking devices in all wires not assigned to carrier service at the points where the cables enter offices in which noise is produced; (2) Attack on noise at points where it is produced by introduction of spark-killers and individual high-frequency choking devices. The lenient noise-to-signal ratio requirement for television mentioned above makes high-frequency television application much easier than message telephone.

While a million-cycle frequency range over more than one pair in an existing cable seems unlikely for telephone message purposes, there are interesting possibilities in the use of lower maximum frequencies. For example, it seems likely that twelve same-directional telephone channels may be obtained from each one of a large fraction of the pairs in existing toll cables and that the crosstalk between the pairs may be kept within proper bounds by simple balancing methods previously described.[3]

## Open Wire

With open wire, conditions are just about the reverse of those with cable. A mile of open wire has an attenuation of only about 1 db at a million cycles as compared to 6 db for the small-sized coaxial. However, overcoming crosstalk between different pairs of wires on a pole line presents very formidable problems while avoiding interference from and to radio systems may be even more formidable.

The attenuation of open-wire pairs has been checked up to several million cycles and it has been found that it behaves as expected. When there is little crosstalk from the high-frequency band to other wires in the lead the attenuation-frequency characteristic is smooth; when there is severe crosstalk, the characteristic is bumpy. For a given length of circuit the attenuation is small compared with that of the small-sized coaxial, and the variation due to changing weather conditions is also small—about one-third that of the coaxial. It is interesting to note, however, that the percentage change in attenuation for the coaxial is less than half that for the open-wire line. While it is, of course, evident that the open-wire transmission variations depend in part on changes in the series resistance of the wires due to changing

temperature and in part to changes in leakage and capacitance due to varying weather conditions, automatic transmission regulating systems similar in general principles to those already developed for other purposes should be adequate to maintain the required stability.

Crosstalk between different circuits becomes so severe at high frequencies that special transposition treatment or respacing of the wires becomes necessary. Minimizing interference from and to radio systems calls for a high degree of balance which may or may not dictate changes in the wire configuration. Here again it is necessary to distinguish between the requirements for television and for message telephone. Tests indicate that, in view of the more lenient television requirements, several million-cycle television channels can be transmitted over different pairs of a single open-wire line without serious disturbance and that to do this it will not be necessary to make radical changes in present wire configurations.

## New Cables

For new construction, if television is not considered, effective carrier telephone systems may be set up by various methods. One might be a very broad-band method, a good example of which is given in the Espenschied-Strieby paper.[1] Another might be a much narrower band method using conductors similar to those in an ordinary cable. In the one case many telephone channels are obtained from a single pair by dividing up a frequency range, say one million cycles wide, into somewhat more than 200 channels. In the other case only 20 odd channels are obtained per pair of wires and use is made of 10 pairs of wires to obtain the same total number. It is too early to say which of these plans might be best under various practical conditions.

To meet future television needs, it may be necessary to provide for transmission of continuous frequency bands a million cycles in width or wider. It is interesting to compare the coaxial with balanced pairs surrounded by individual shields for such transmission.

For 6 db loss per mile at a million cycles, it works out that a solid copper coaxial unit with the rubber disc insulation described in the Espenschied-Strieby paper [1] has an internal diameter (inside of shield) of about 0.25 inch. For the same attenuation a pair of wires, each the same size as the central wire in the coaxial unit (70 mils diameter), insulated with rubber discs and with a copper shield, will have an inside diameter under the shield of about 0.4 inch. For outside conductors or shields made of lead the inside diameters become about 0.4 inch for the coaxial and 0.5 inch for the balanced pair. Therefore, if the thickness of the outer conductor is determined by mechanical considerations

rather than outside interference, the coaxial is smaller and cheaper. As the frequency is made higher, the shielding from outside interference afforded by the surrounding cylinder increases, so that at very high frequencies mechanical considerations alone control and the coaxial is clearly cheaper than the balanced shielded pair.

In the frequency range up to about a million cycles, however, interference from outside sources, including natural static and radio, must be considered in determining the thickness of the surrounding cylinder and the cost comparison is not so clear. It will be evident that as regards shielding, the balanced pair is at a large advantage because the two sides of the circuit are designed to be electrically similar. By proper care in manufacture this balance can readily be made sufficient to insure adequate shielding with a surrounding lead tube of thickness determined solely by mechanical considerations.

With a coaxial structure, however, it appears likely that to keep interference within proper bounds, a simple lead tube must be made considerably thicker than required by mechanical considerations, so that such a structure would probably be more expensive than a lead shielded pair. However, by adding other materials an adequately shielded coaxial unit can be constructed which will have a considerably thinner outside wall.

For example, there is described in the Espenschied-Strieby paper [1] a coaxial unit in which the inside diameter is minimized by first using copper tapes, the thickness of wall is minimized by adding thin iron tapes and the whole is made waterproof by a thin surrounding lead tube. This results in a unit of smaller inside and outside diameters than those of a lead tube surrounding a balanced pair of like attenuation. Since, however, the wall of the coaxial is thicker and the structure more complicated, the costs of the two units are estimated to be not greatly different when they are designed for the frequency range up to a million cycles. A minor advantage for the balanced pair remains, however, in that, whatever may be the top frequency, there is no limitation as to the lowest frequency permissible for interference reasons.

In the above discussion of new cable construction the amplifiers and transmission regulators required have not been mentioned. If similar conducting and insulating materials are used, shielded balanced pairs and coaxials have similar transmission-frequency characteristics. The variations with temperature are also similar. The factors which limit the overall amplifications are also the same. There is only one important point of difference between the amplifiers required for the two systems. This is the necessity for input and output transformers to be

balanced to ground with the balanced pairs. The excellence of balance required, of course, depends on the extent to which balance is relied on to reduce the required thickness of sheath. In view of the fact that very thin sheaths are impracticable for mechanical reasons it appears probable that only very modest requirements as to balance need be imposed on the design of these transformers.

To provide several circuits in new cables for meeting wide-band television needs, another method may be considered, that is, to provide balanced pairs considerably larger in size than ordinary pairs and with the rubber disc form of construction, or other form giving low dielectric losses, but with no shields at all around the individual pairs. Shielding from outside disturbances would be adequately provided by the outside lead sheath of the cable. Crosstalk between different pairs would be the principal concern. If all of the high-frequency pairs were to be used for television transmission, the crosstalk requirements, as already mentioned, would not be severe, so that by careful design the crosstalk could readily be kept within proper bounds—of course, restricting transmission of all wide-frequency bands to a single direction within a single sheath. Such high-frequency balanced pairs might prove suitable for telephone message circuits also. If not, the high-frequency pairs would be restricted to television only, and other pairs, worked at lower frequencies, would be provided for telephone message service.

## Summary

It appears feasible under certain conditions to transmit continuous frequency ranges of 1,000,000 cycles or more over conductors in the existing telephone plant. This may some day prove very important, particularly if the art of television develops to the point of calling for such wide frequency range circuits to carry television impulses around the country as sound programs are now carried.

For new construction, the balanced type of circuit, as well as the unbalanced coaxial circuit, offers many interesting possibilities.

# A Survey of Magnetic Materials in Relation to Structure*

By W. C. ELLIS and EARLE E. SCHUMACHER

The structure dependence of magnetic characteristics of the more important magnetic materials is discussed. The natural grouping into soft magnetic materials and hard or permanent magnet materials is used. In the first group are: magnetic iron, silicon steel and the alloys of iron, nickel and cobalt. The latter group includes the well-known carbon and alloy steels, and also the newly developed precipitation hardening $\alpha$-solid solution types. Examples of the latter are the iron-nickel-aluminum and the iron-cobalt-molybdenum alloys.

In discussing the properties, the relationship to structure is emphasized. With the soft magnetic materials purification and the development of proper structural configuration are important. The special metallurgical control required to produce certain magnetic materials in dust form is described. The generalization is made that permanent magnet characteristics are due to precipitation effects resulting from decomposition of supersaturated solid solutions.

## Introduction

MAGNETIC materials may be classified according to properties into two groups: (1) soft magnetic materials, and (2) hard or permanent magnet materials. The differences in properties of the two groups are illustrated in Fig. 1, which shows typical magnetization cycles for the two types. The dashed line is a normal magnetization curve while the closed full line is known as the hysteresis loop.

A soft magnetic material is, in general, characterized by a steeply ascending magnetization curve; that is, large values of flux density are produced by small magnetizing forces. For certain applications where the flux density is low, the initial portion of the curve is important. For intermediate flux density applications the steeply ascending portion is of paramount interest while for higher densities, the upper portion is of prime importance.

Another term which is widely used in discussing soft magnetic materials is permeability. The permeability at a given flux density is the slope of the straight line joining that point on the magnetization curve with the origin. Algebraically, it is the ratio of the flux density, $B$, to the magnetizing force, $H$. It varies with the flux density, and for soft magnetic materials at low magnetizing forces, is a large quantity. The permeabilities usually reported in discussing magnetic properties are the initial permeability, $\mu_0$, and the maximum permeability, $\mu_{max.}$. The initial permeability is the initial slope of the magnetization curve. The

* Published in *Metals and Alloys*, first part in December 1934 issue, second and concluding part in January 1935 issue.

maximum permeability is the maximum value that the ratio $B/H$ attains.

Hysteresis is also of importance, especially in alternating current applications. The phenomenon of magnetic hysteresis results in a loss of energy in a magnetic material when the material is carried through a magnetization cycle. The loss of energy is proportional to the area of the hysteresis loop. In soft magnetic materials, the aim, in practically all cases, is to keep this loss a minimum.

A hard or permanent magnet material is characterized by a gradually ascending magnetization curve; that is, the material at all magnet-

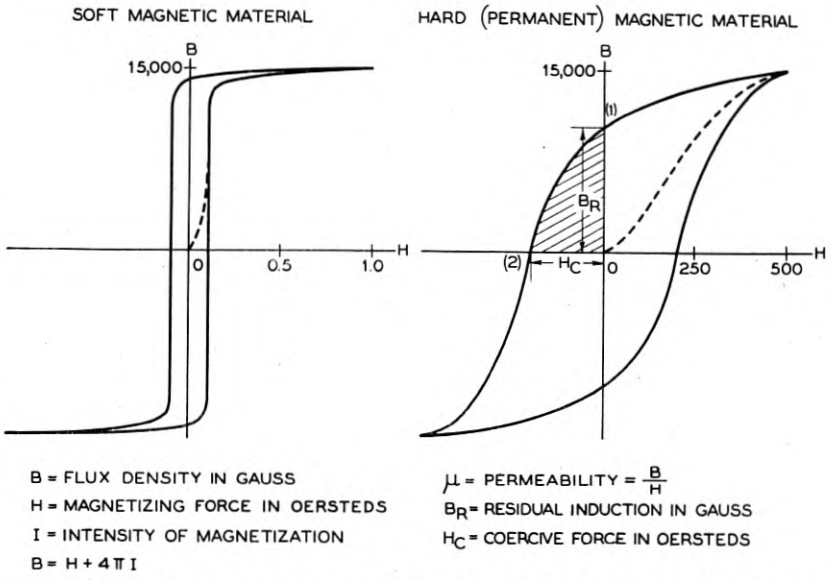


Fig. 1—Schematic representation comparing characteristics of soft and permanent magnet materials.

izing forces has a relatively low permeability. To attain practical saturation is difficult, requiring in some alloys magnetizing forces of thousands of oersteds. The important properties are associated with the demagnetization curve, that portion of the hysteresis loop between points (1) and (2) in Fig. 1. In general the hysteresis loop in its entirety is of little importance. The constants usually determined in permanent magnet investigations are the residual induction, $B_r$, and the coercive force, $H_c$. The product of these two quantities ($B_r H_c$), which is roughly proportional to $(BH)_{max.}$ on the demagnetization curve, has been considered by some [1] as a quality index for permanent magnet materials.

Summarizing the general properties of soft and of hard magnetic materials, in the former, high maximum permeability is associated with low coercive force and a hysteresis loop of small area. In the latter, a low permeability is associated with a high coercive force and a large hysteresis loop. The range of properties which can be obtained in magnetic materials by alloying and other metallurgical control is most remarkable. Permeabilities from 1 to over 600,000 and coercive forces from 0.012 to 600 oersteds, are attainable.

In manufacturing soft magnetic materials, the metallurgist strives constantly to eliminate those chemical elements (impurities) which broaden the hysteresis loop, while in the case of the permanent magnet materials he intentionally adds certain of the same, or other elements, and resorts to heat treatments which broaden the loop. Some details of these procedures with well known materials, and also with some more recently developed, are described in the following pages.

## Soft Magnetic Materials

### *Magnetic Iron*

Of the ferromagnetic elements iron, nickel, and cobalt, iron, at the present time, is the only one of industrial importance as a soft magnetic material in the unalloyed condition. Throughout the last half century, a continual improvement in the magnetic quality of iron has been effected until at present, laboratory samples have been prepared with permeabilities [2, 11] of over 200,000. Careful analysis of the data shows that this improvement has paralleled closely increases in the purity of the iron.

Since the production of iron of high magnetic quality is of considerable interest at this time, it seems worth while to examine, in some detail, the methods which have been used for producing high purity iron. Methods which have been developed for the purification of iron, have, in each instance, removed from the iron, elements (impurities) which are harmful to easy magnetization.

The elements which enter into iron as impurities may be grouped, structurally, into two classes: substitutional elements and interstitial elements. The substitutional elements exist in the iron structure at lattice points; that is, they substitute for iron atoms in the lattice shown in Fig. 2. These elements include most of the metals, for example: nickel, cobalt, manganese and silicon. The interstitial elements enter into the iron lattice at intermediate points; that is, they take positions in the structure between the iron atoms. Elements included in this classification are carbon, oxygen, and nitrogen. It is these latter elements which, although present in small quantities, cause severe

strains in the lattice, and are believed, on this account, to be particularly harmful to the magnetic softness.

The solubilities of the interstitial elements at room temperature have been determined by a number of investigators. The most reliable values, which, because of the difficulties involved, probably should be considered tentative, place the room temperature solubility of carbon at approximately 0.008 per cent;[3] the solubility of oxygen at 0.01 per cent;[4] of nitrogen at 0.015 per cent;[5] and of sulfur at 0.015 per cent.[6]

The solubilities of the interstitial elements vary with the temperature, increasing, in general, at higher temperatures. Varying solubility provides the possibility of precipitation hardening, which, in the case of iron, results in magnetic hardening. It is this latter type of

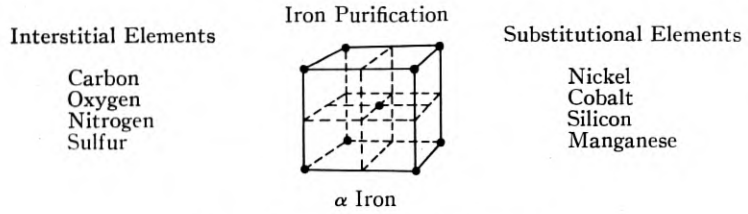


Fig. 2—Schematic representation of methods of purifying iron for magnetic purposes. The lattice structure shown is that for α-iron. The iron atoms are located at the cube corners and at the center and are represented by the circles shown.

hardening to which the magnetic aging of iron is attributed. This is possible because the supersaturated phase will in most of these systems precipitate the interstitial element at room temperature in a form which produces severe strains with deleterious effects on magnetic softness. Because of their harmful effects, the purification of iron for soft magnetic purposes has been in the direction of eliminating these interstitial elements. The three methods of purification which have been used are illustrated in the chart shown in Fig. 2.

The earliest method used was the addition in the molten state of a substitutional element which diminishes the harmful effect of interstitial elements in the iron lattice. Silicon* is an element beneficial in

* Messrs. Barrett, Brown, and Hadfield [7] reported the magnetic properties of iron-silicon alloys in 1900. More recent studies have been reported by Yensen.[7]

this respect. Its action may be explained by postulating that its presence decreases the solubility of carbon in iron. The silicon enters the iron lattice as a substitutional element, and as such, is less harmful to the magnetic softness than the interstitial element, carbon, which is removed. Since silicon is also a strong deoxidizer, it will remove oxygen from the iron. These reactions do not completely eliminate the carbon and oxygen, and consequently, although a noticeable increase in magnetic softness is obtained, the high values of permeability achieved with the more efficient methods of purification are not

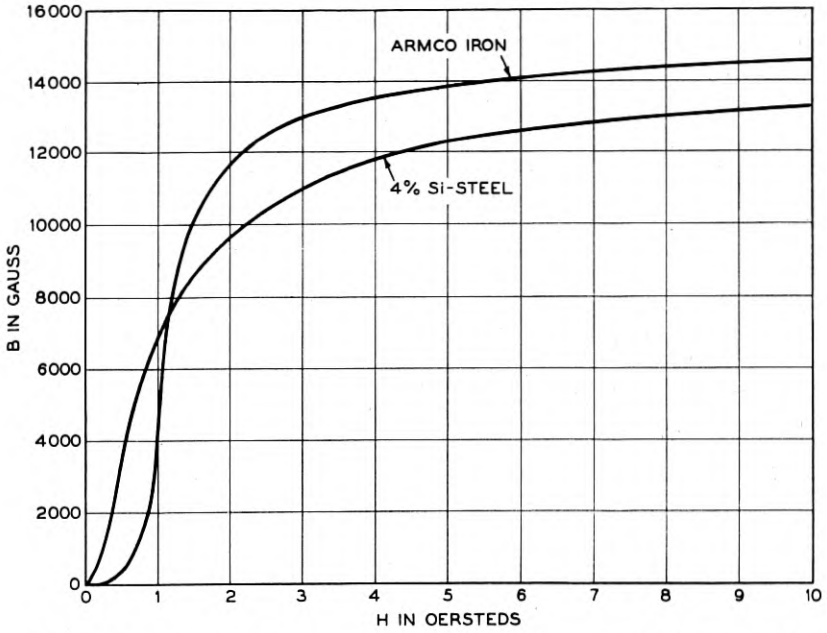


Fig. 3—A comparison of the magnetization curve of commercial 4 per cent silicon steel[8] with that of Armco iron[17] showing the superiority of the former at low induction densities.

reached. The improvement attained by this treatment is illustrated in Fig. 3 by comparison of typical magnetization curves for armco iron [17] and for a commercial 4 per cent silicon steel.[8]

The second method of producing a purer iron consists of eliminating some of the harmful elements by melting in a high vacuum.[9] Under these conditions, the residual carbon and oxygen in the iron will combine to form carbon monoxide, which is removed by the exhausting pumps. At the same time, due to the dissociation of the nitrides at the high temperature, the nitrogen is similarly removed. Again, the

reactions do not go to completion, and residual impurities remain. Another obvious disadvantage is the fact that it would be only accidental if oxygen and carbon were present in the iron in the right proportions to eliminate one another. If the oxygen and carbon contents of the raw materials are known, adjustments can theoretically be made by adding either carbon or iron oxide, but if such a process is in commercial operation, the writers are not aware of it.

Yensen [9] reports a laboratory experiment in which this method of purification was used. A number of samples were prepared in high vacuum from electrolytic iron. To successive samples, increasing amounts of carbon were added to a maximum of 0.5 per cent. The maximum permeability increased with additions of carbon from approximately 25,000 for the sample with no addition to a maximum of 61,000 for a 0.06 per cent addition. From this value, the permeability dropped rapidly with further additions of carbon. The maximum value of 61,000 is explained as due to the fortuitous coincidence that the correct amounts of carbon and oxygen were present for optimum elimination.

Vacuum melting and vacuum purification present some interesting possibilities in connection with the preparation of magnetic materials. Commercial vacuum melting has been developed in Germany at the Heraeus plant,[10] where furnaces with capacities of 5 tons are in operation. It is stated that capacities of 20 to 30 tons can be achieved without large departures from the present designs. Vacuum melted metal is mechanically softer and works more readily than metal prepared by ordinary melting procedures. For magnetic alloys, the combination of vacuum melting and casting provides facilities for preparing material without contamination by the atmosphere. There seems to be no reason why one cannot go further and actually carry on refining operations in the furnace. The future of vacuum melting and casting appears extremely rich in possibilities, particularly for producing magnetic alloys where high purity is a primary consideration.

The third method of purification of iron is by treatment with an element which reacts with the interstitial elements to form gases which are removed. Hydrogen has been used for this purpose by Cioffi,[11] and the method has been described recently. The treatment is carried out at high temperatures between 1300° C. and the melting point of iron. Cioffi states that ordinary amounts of carbon, oxygen, nitrogen, sulfur and phosphorus are reduced to very small quantities by hydrogen treatment. The excess hydrogen is either liberated as the metal cools, or, if it remains in the metal, is without harmful effects on the magnetic characteristics. A value of maximum permeability of 280,000

has been reported [11] for iron prepared in this manner. In Fig. 4 is reproduced a curve from the unpublished work of Cioffi showing the enormous improvement in magnetic softness achieved by the hydrogen treatment.

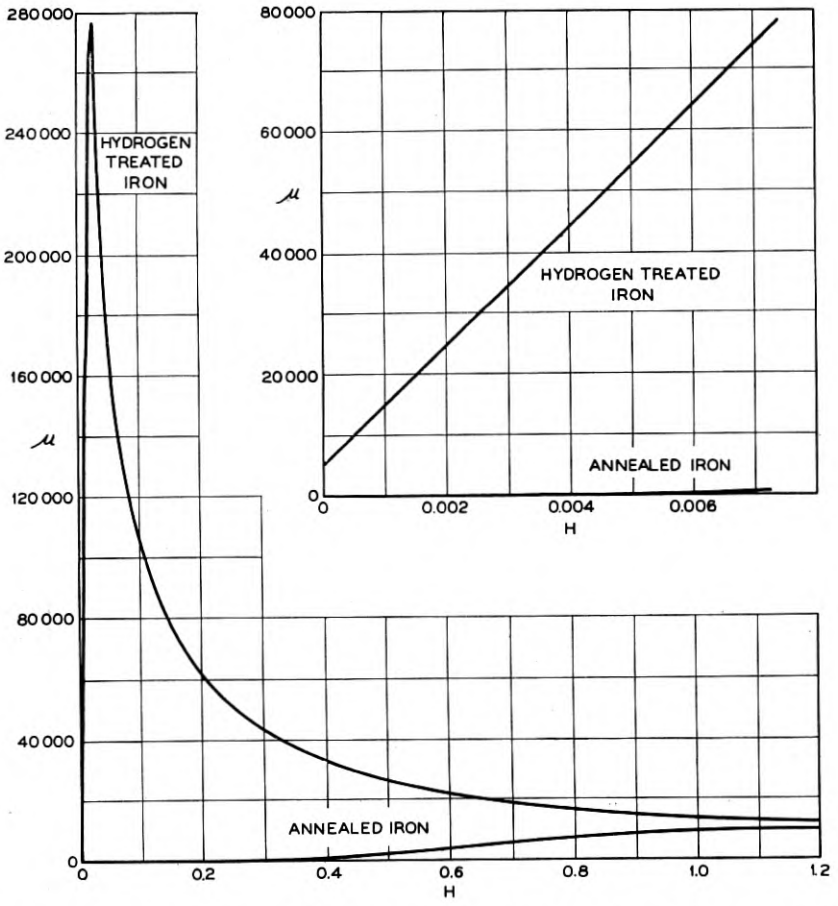


Fig. 4—The permeability of a laboratory sample of hydrogen treated iron compared with that of untreated iron—Cioffi[11].

The mechanism of the purification is outlined by the reactions in Fig. 2. The carbon is eliminated as a hydrocarbon gas, probably methane; the oxygen as water vapor; the nitrogen as ammonia or as nitrogen gas due to thermal dissociation of nitrides; the sulfur as hydrogen sulfide; and the phosphorus probably as phosphine. The most probable explanation to account for the purification is that, at the

high temperatures, the impurities continually diffuse to the surface of the sample where the reaction with hydrogen occurs. Another possibility is that the reactions occur within the body of the sample, in which case the hydrogen must diffuse in and the reaction products out. The first possibility appears the more promising in that the diffusion of reaction products through the iron samples would be difficult, since, in general, the products are rather large molecules.

Another type of high purity iron which is being used for magnetic purposes is carbonyl iron [12] manufactured by the I. G. Farbenindustrie in Germany. It is prepared by first forming iron carbonyl under suitable conditions of temperature and pressure, and subsequently decomposing the carbonyl. For the best quality, the material is further purified by a low temperature hydrogen treatment. This iron is in the form of spherical particles, a few microns in diameter.

One use for this material is in cores of high frequency inductance coils. The cores of such coils are prepared by pressing the insulated magnetic particles into the desired core shape.

The material is also used in preparing iron sheet [13] and as an alloying constituent for the manufacture of iron-nickel alloy sheet.[14] The process of alloying consists of pressing the powders and sintering at a high temperature, followed by working and annealing in hydrogen or vacuum. Typical permeability values for iron sheet prepared in this manner are: $\mu_0 = 2000$–$3000$; $\mu_{max.} = 15{,}000$. The values are decidedly lower than those for a laboratory sample of hydrogen-treated iron shown in Fig. 4.

### *Iron Nickel Alloys*

No alloy system has been more fruitful in yielding interesting and useful magnetic alloys than that of iron and nickel. In this system, the three regions marked with arrows on the constitutional diagram in Fig. 5 are of principal importance.

In the proximity of 25 per cent of nickel, which is the region of retarded phase change, alloys can be obtained at room temperature in the non-ferromagnetic state. How this is accomplished is evident from a consideration of the constitutional diagram. The $\gamma$ solid solution above the magnetic transformation is non-ferromagnetic while the $\alpha$ solid solution is ferromagnetic. When alloys in the region of 25 per cent nickel are cooled, the transformation to the ferromagnetic state is not completed until temperatures below room temperature are reached. If the alloy is cooled until the ferromagnetic structure is obtained, the transformation to the non-ferromagnetic state does not occur on heating until the temperature has reached approximately 600° C.

Elmen [17] has investigated alloys containing more than 30 per cent of nickel which are characterized by great magnetic softness and have been called "permalloys," [17] the name suggesting high permeability. These alloys in the region of 40–55 per cent nickel find industrial application as soft magnetic materials. They are used in telephone and radio transformers and telephone relays. An alloy typical of this region is the one containing 45 per cent of nickel. A curve [15] showing

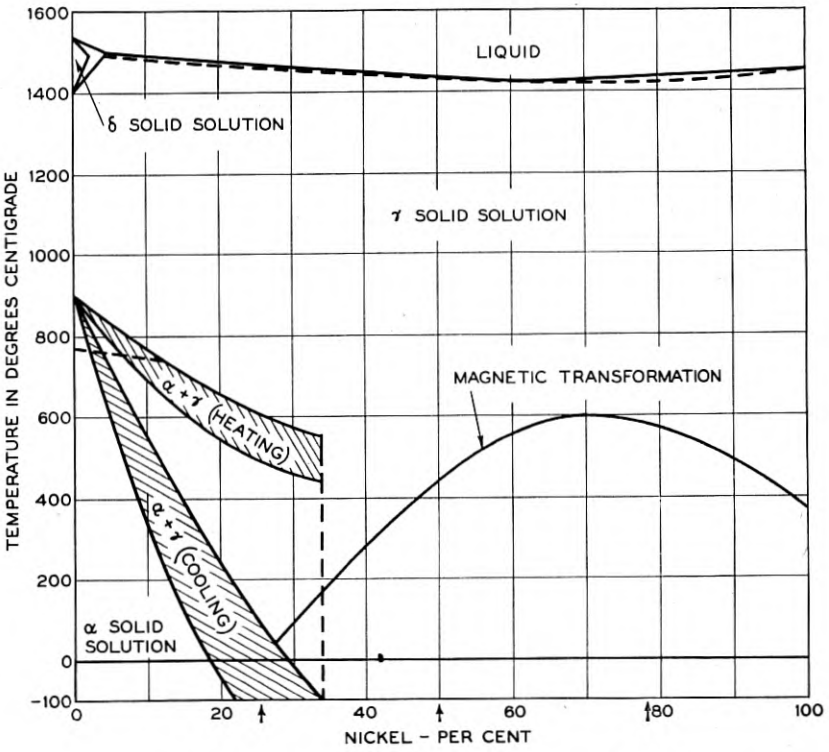


Fig. 5—The constitutional diagram of iron and nickel.

the permeability of this alloy is plotted in Fig. 6. From the data given on this curve, the material exhibits an initial permeability of 2000 and a maximum permeability of 16,000.

Another alloy in this region is the one containing 50 per cent of nickel which has been investigated by Yensen. He has applied the name "Hypernik" [16] to it. Yensen, by giving this material a special heat treatment, obtained an initial permeability of 4000 and maximum permeability of 70,000.

Arnold and Elmen [17] have investigated in great detail the alloys in the region of 70 to 80 per cent nickel. These alloys are used where extremely high initial permeability is required. One especially interesting use is in the loading of submarine telegraph cable. An alloy typical of this region is the one containing 78.5 per cent of nickel. Magnetization curves for this alloy after two different heat-treatments are compared with the curve for annealed armco iron in Fig. 7. In the

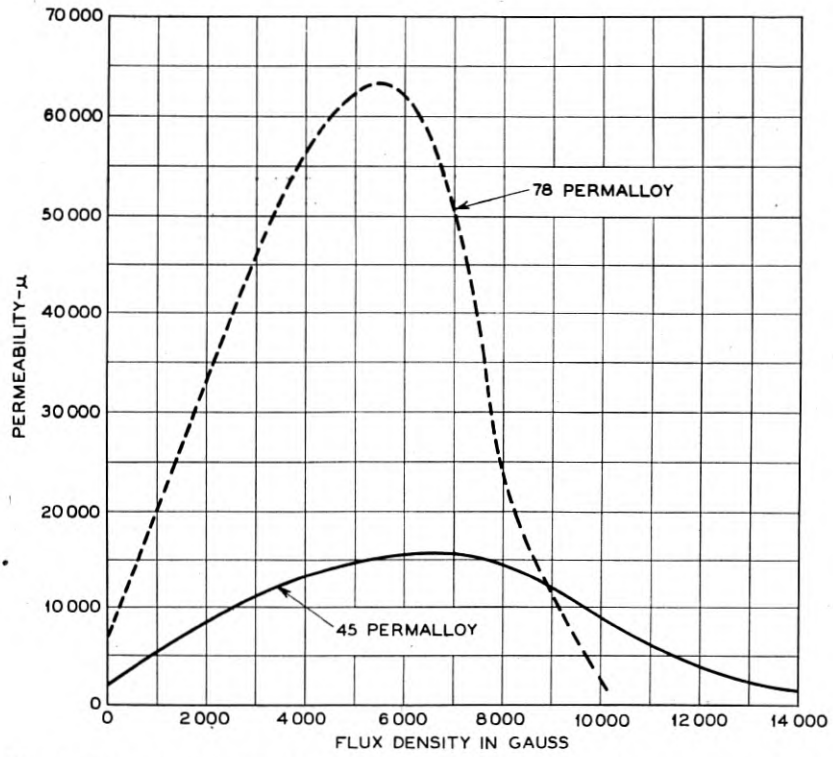


Fig. 6—The permeability of 45 permalloy compared with that of 78 permalloy—Pettit[15].

air-quenched condition, in which the alloy is magnetically the softer, typical values for the initial and maximum permeability are 7000 and 65,000, respectively.

In comparing the properties of 45 per cent nickel and 78 per cent nickel alloys with reference to Fig. 6, the initial permeability is higher for the higher nickel alloy. Therefore, for very low magnetizing forces, the higher nickel alloy is characterized by higher flux densities. For higher magnetizing forces, the 45 per cent alloy has more attractive

properties in that the flux density is higher. In addition, the electrical resistivity of the alloy is approximately three times that of the 78 per cent alloy. The electrical resistivity is important in alternating current applications since the eddy current loss in sheet of the same physical dimensions varies inversely with the electrical resistivity of the material.

To the metallurgist the effects of heat treatment in establishing the magnetic properties of iron-nickel alloys possess a great fascination.

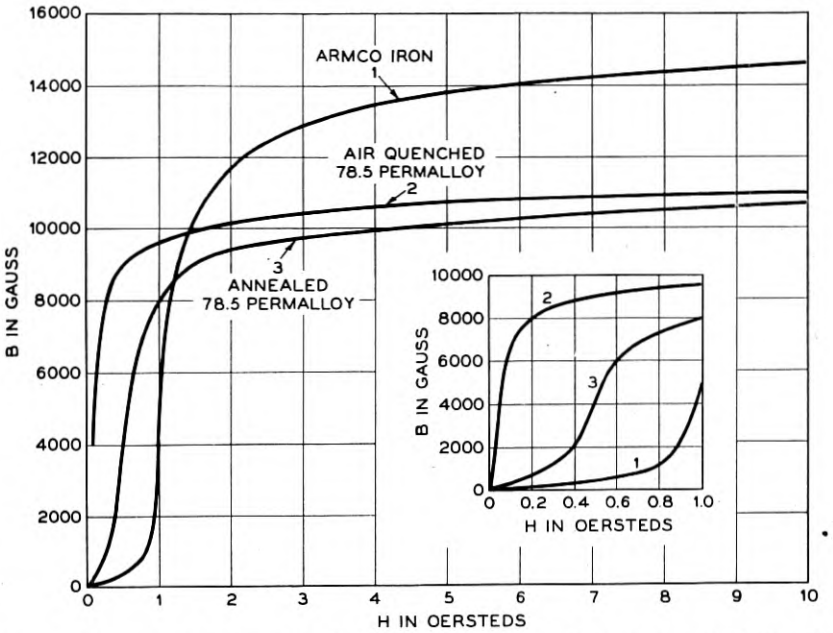


Fig. 7—Magnetization curves for 78.5 permalloy after two heat treatments—Elmen[17].

The effects in question are illustrated by curves shown in Fig. 8. In the region of 78 per cent nickel, the maximum magnetic softness is exhibited after air-quenching, while increases in magnetic hardness are obtained after annealing or baking for extended times at low temperatures. The compositions near 50 per cent nickel are not as much affected by heat treatment.

Elmen [17] attributed the heat treating effect to the separation of the homogeneous phase, stable at high temperatures, into two phases on slow cooling. The two phase structure under these conditions would be less magnetic. More recently in Germany, investigators [18] have advanced the analysis by postulating that in the region of 76 per cent nickel, where the ratio of iron to nickel atoms is 1 : 3, the alloy can exist

in two states: one, a solid solution state in which the atoms of iron and nickel are in a disordered distribution; and two, a state in which the iron and nickel atoms are regularly distributed.

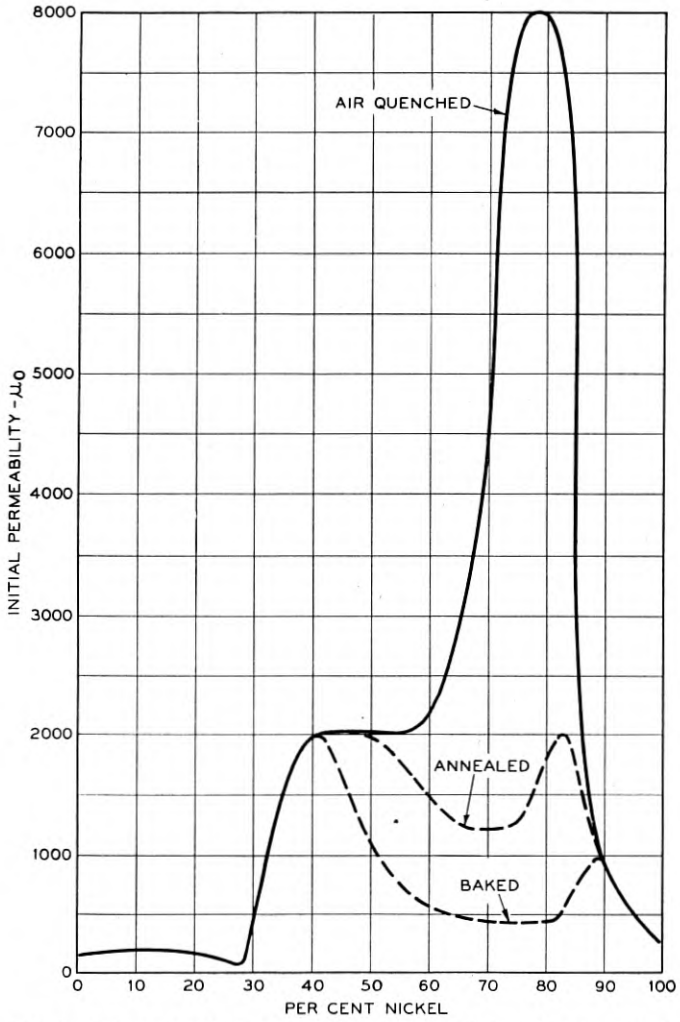


Fig. 8—Effect of heat treatment on initial permeabilities of iron-nickel alloys—Elmen[17].

In the disordered distribution the lattice points are fixed but the atom positions are not; that is, any lattice point may be occupied by either an iron or nickel atom. In the regular or ordered distribution, the lattice points are fixed and also the atom positions. The second

arrangement may be clearer to some if it is considered a "compound," although strictly speaking, the arrangement does not correspond to a compound in the chemical sense. The regular arrangement is now ordinarily called a "superlattice" or a "superstructure," and as such, will be designated in this paper. If it is clearer to the reader, he may substitute for the term "superlattice" or "superstructure," the term "compound" without great sacrifice in the sense of the discussion.

With this picture in mind, the effect of heat treatment is immediately evident. Rapid cooling prevents the formation of the "superlattice" arrangement, which occurs somewhere in the region of 500° C., and a disordered distribution of atoms is retained. This is the distribution possessing the higher magnetic permeabilities. Slow cooling or baking promotes the transformation to a superlattice of lower magnetic quality.

Heat treatment effects are much less in the alloys in the region of 50 per cent nickel, and in fact would not be expected, since a face centered cubic binary alloy with the atomic ratio of 1 : 1 would not be expected to have a special structure (superstructure).

The hypothesis of superlattice formation in the 75 per cent nickel region is supported further by the effect of heat treatment on other physical characteristics, for example the electrical resistivity and the tensile strength. That such special structures are formed in the solid states is well established from detailed studies of the copper-gold system.[19] Unfortunately, in the nickel-iron system where the nickel and iron atoms are so near in atomic number, the detection of superstructure by x-ray methods appears, at the present time, a difficult task.

The effect of heat treatment and the general magnetic softness of iron-nickel alloys in the region of 75 per cent nickel may be explained by another hypothesis which returns for its basis to the considerations set forth in the section on high purity iron. The improvement in magnetic softness may be attributed, first, to the effect of nickel in decreasing the residual quantities of interstitial elements originally present in iron. Whether or not such an effect is present awaits a careful investigation of the interstitial element content of iron-nickel alloys. Secondly, assuming the interstitial elements the responsible factors, the effect [20] of heat treatment on magnetic quality can be explained in that the quench (rapid cooling) retains the residual elements in solution, while slow cooling permits precipitation in a form more deleterious to magnetic quality.

The two hypotheses possess interesting possibilities, but it appears that further investigation will be required before the structural relationships are definitely established.

The permalloys in the region of 75 per cent nickel have been further modified by the addition of third elements [17] which are non-ferromagnetic. Molybdenum and chromium are elements of this nature which confer on the alloys certain characteristics of sufficient interest to warrant discussion. The effects of molybdenum and chromium on the magnetic quality are shown in Fig. 9, which indicates that higher initial permeabilities may be expected with the ternary alloys. The addition of a third element increases the electrical resistivity of the alloy. This is of importance in alternating current applications where eddy current losses are a consideration. In commercial manufacture the third element is particularly valuable, since it decreases the sensi-

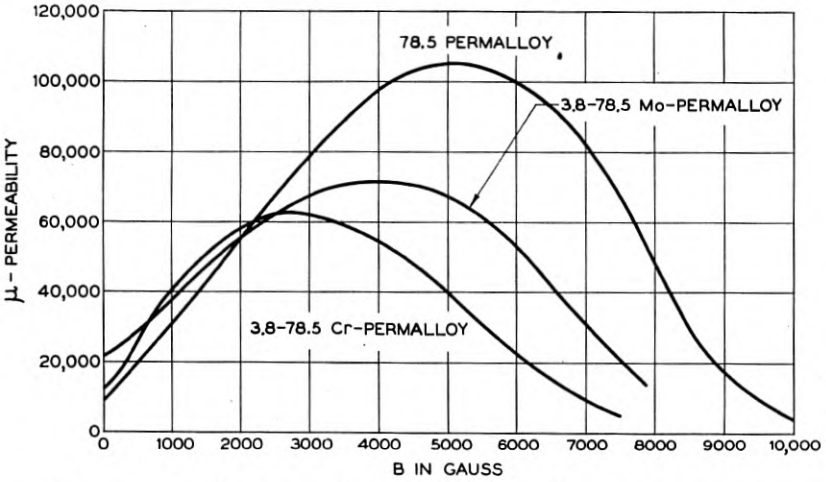


Fig. 9—Permeability curves obtained for laboratories' samples comparing the properties of chromium permalloy and molybdenum permalloy with those of standard permalloy—Elmen [17].

tivity of the magnetic properties of the alloy to the rate of cooling after annealing and, in fact, permits the attaining of high permeabilities with slow cooling. The significance of this from a structural viewpoint is that the third element keeps the iron and nickel atoms in a disordered distribution.

Iron-nickel alloys have a limited field of application in the form of thin tape in the hard rolled condition. A representative material is one containing from 30 to 70 per cent nickel and, in addition, a few per cent of aluminum or copper, which in the hard rolled condition has been called "isoperm." [21] The material is reported to have a low permeability of 50 to 75, which is fairly constant over a range of magnetizing forces of 0–10 oersteds. Laminated cores of this material have prop-

erties similar, but somewhat inferior to, those of magnetic dust cores [27] which are described in a later section of the paper.

*Iron-Cobalt Alloys*

Iron-cobalt alloys in the region of 50 per cent cobalt are of importance for magnetic purposes because of the high values of flux densities obtained with medium magnetizing forces. The desirable magnetic properties of the alloy containing 50 per cent cobalt were first reported by one [22] of the authors. A very complete investigation of alloys over the whole range of cobalt contents has been described by Elmen [17] from whose paper the curves shown in Fig. 10 are taken. With

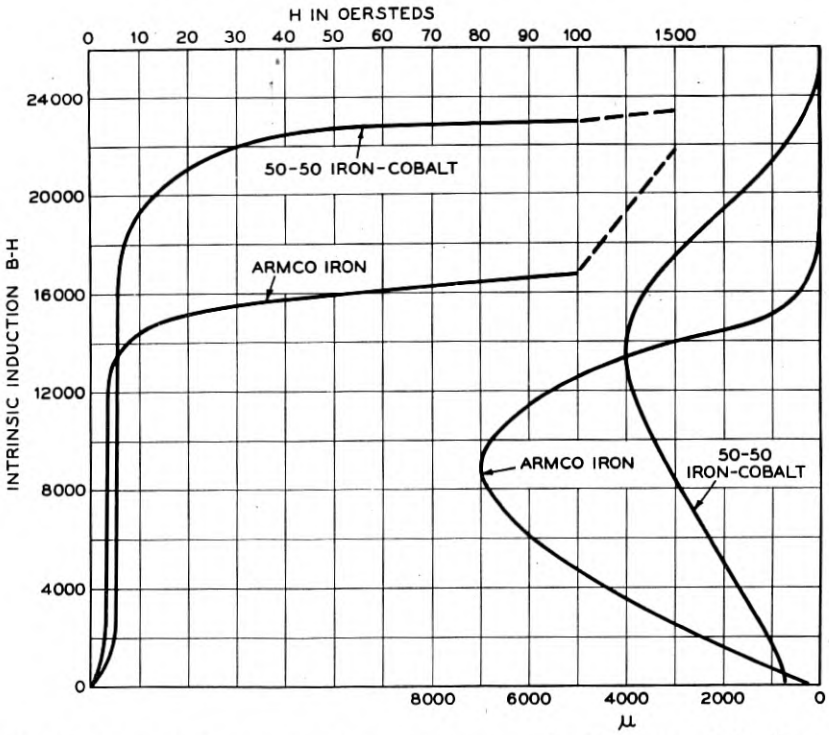


Fig. 10—Magnetization curve and permeability curve for 50–50 iron-cobalt alloy—Elmen[17].

magnetizing forces from 5 to over 1500 oersteds the intrinsic induction is greater than that obtained with iron.

Because of anomalies in properties of the alloys in the region of 50 per cent cobalt, the structural features of the alloys become of interest. In addition to the maximum in intrinsic induction, the electrical con-

ductivity in this region at room temperature is greater than that of either of the components. The anomalous behavior was attributed several years ago by one of the authors,[22] not to compound formation, but to a "peculiar solid solution." Kussman, Scharnow and Schulze [22] in a recent paper attribute the effects to the formation of superstructure in the body centered lattice which exists at this point. This, in a sense, is an extension of the earlier explanation. Here again, it appears that further work is warranted and is required before the structure is completely understood.

In connection with the metallurgy of the 50–50 iron-cobalt alloy, there are some interesting considerations. The binary alloy can be worked hot, but is extremely brittle when cold. This imposes definite limitations where thin sheet produced by cold rolling is required. The limitation can be overcome by adding a few percent of vanadium [23] to the composition. This alloy can be worked hot, and after a quench from a high temperature, also can be cold-rolled to thin sheet. The magnetic characteristics are not greatly affected by the vanadium, provided the addition is small.

The structural changes resulting from the additions of vanadium are not clearly understood. It is known that vanadium retards the transformation which occurs in these alloys at approximately 900° C. and therein lies a possible explanation for its action. The high temperature modification is malleable as shown by satisfactory hot working properties. By quenching, sufficient of this modification may be preserved at room temperature, when vanadium is present, to permit cold rolling.

### *Iron-Cobalt-Nickel Alloys*

In the ternary system, iron-cobalt-nickel, is a region in which the alloys exhibit an unusual and useful property, namely, constant permeability in low fields. Because of this characteristic, alloys in this region have been named perminvars.[24] A typical alloy with marked perminvar characteristics contains 45 per cent nickel, 25 per cent cobalt and 30 per cent iron. The effects of heat treatment are illustrated in Fig. 11. The effect on the magnetization curve of air quenching, annealing, and baking at a low temperature is shown for this typical alloy. The extension of constant permeability to higher magnetizing forces by a low temperature bake is clearly demonstrated.

The constancy of permeability in perminvar has been explained in one hypothesis as due to the presence of two constituents in the alloy, one a soft magnetic material, the other a hard magnetic material. The experimental evidence for this hypothesis is derived largely from the constricted hysteresis loops which are obtained for these alloys after a

baking treatment. A similar type of constricted loop can be obtained in a magnetic sample which is composed of thin parallel sheets of a soft and of a hard magnetic material magnetized longitudinally.

It may be that the two constituents result from a process similar to that described for the iron-nickel alloys in the region of 75 per cent

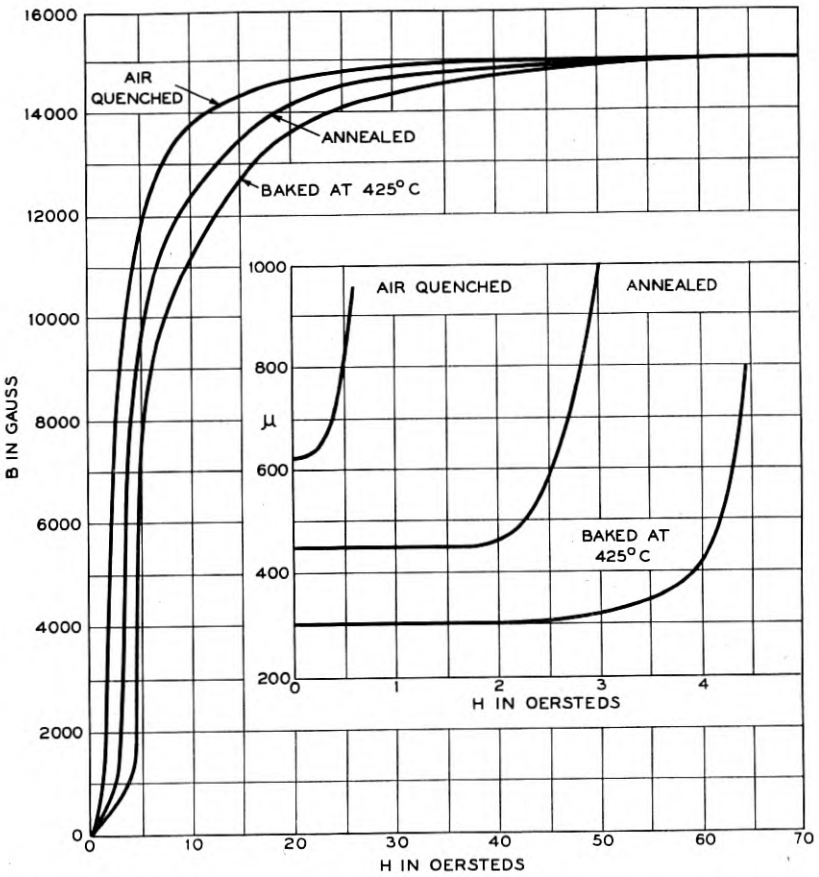


Fig. 11—Magnetization curves for a perminvar containing 45 per cent nickel, 25 per cent cobalt and 30 per cent iron after several heat treatments—Elmen [17].

nickel. In line with this, the baking treatment may form a superstructure [24] which is magnetically hard, but to a greater degree, than the one in the binary iron-nickel alloys. It is probable that the similar effects of heat treatments in the binary and ternary alloys are due to similar structural transformations.

*Heat Treatment of Magnetic Material in Magnetic Field* [25]

Recently it has been reported that radical changes have been produced in the magnetic properties of some ferro-magnetic alloys by heat treatment in a magnetic field. For example, the maximum permeability of the permalloy containing 65 per cent of nickel has been in-

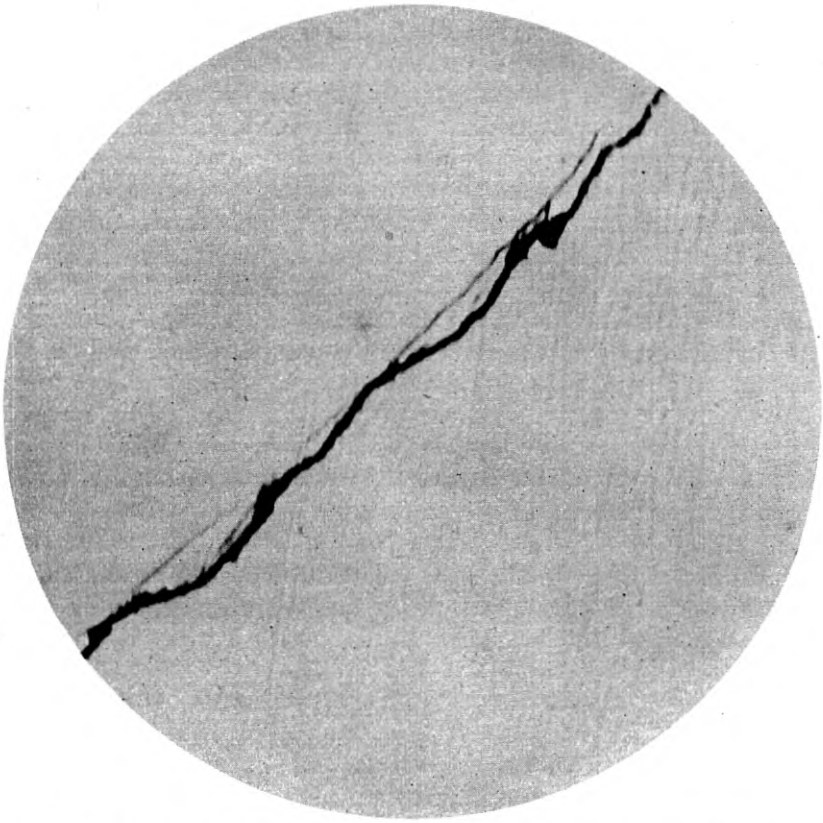

Fig. 12—Structure of an iron-nickel alloy, as cast, containing 80 per cent nickel and 0.030 per cent sulphur. The sulfide constituent exists as a brittle film surrounding the crystallites. Mag. 2000 ×.

creased from 10,000 to 250,000 by heating to 700° C. and cooling slowly in a field of 15 oersteds. The maximum permeability has been further increased to 600,000 by so treating a specimen which had previously been treated in hydrogen at 1400° C. This value of maximum permeability is the highest which has been reported for any ferromagnetic material.

In the iron-cobalt-nickel alloys, high values of maximum permeability are also generally obtained by such a heat treatment, provided the magnetic change point is higher than about 500° C. Another feature in connection with this special heat treatment is that the heat-treated specimens have hysteresis loops which tend to have sharp corners and vertical sides.

An interesting explanation for the effect of heat treatment in a magnetic field is given by Bozorth.[25] According to modern magnetic theory, when a specimen is cooled through the Curie point in the absence of an applied field, small regions or domains are magnetized to saturation in definite directions. The specimen in its entirety exhibits no external magnetic effects since the vector sum of the individual effects would be nearly zero and not detectable. The magnetization of the domains produces internal strains in the material due to lengthening in one direction and contraction at right angles (magnetostriction). By the application of an external magnetic field in the temperature region near to, but below, the Curie point, (1) the magnetic domains are oriented nearly in the direction of the applied field, and (2) the internal strains which were introduced by the local magnetization at the Curie point are removed by plastic flow of the surrounding domains. For this to occur, the temperature of Curie point must be higher than that at which plastic flow begins. Subsequent application of an external magnetic field at room temperature in the direction of, or opposite to, the direction of the magnetization at the higher temperature introduces no strains in the material tending to decrease the ease of magnetization. For that reason relatively higher permeabilities are obtained than with material heat treated in the absence of a field, in which case opposing strains are present.

The theory also offers another explanation for the effects of heat treatment on the permalloys in the region of 75 per cent nickel. This explanation will not be discussed further here except to say that it is concerned with the same magnetostrictive strains as have been mentioned above.

The fundamental idea of the theory proposed to account for the effects of heat treatment in a magnetic field, and the general effects of heat treatment on certain ferromagnetic materials, is that the magnetostriction associated with the local magnetization in the domains is sufficient to cause plastic flow. Experimental data indicate that the magnetostrictive stress is sufficient to cause such plastic flow at about the temperature which is known to be critical for the heat treatment of the permalloys.

### *Heusler Alloys*

One of the most interesting groups of magnetic materials comprises the Heusler alloys [26] which are ferromagnetic although composed of non-ferromagnetic elements. A representative alloy of this type is one containing copper, manganese, and aluminum. The magnetic properties are associated with the composition $Cu_2MnAl$, since the magnetic intensity of the alloys varies in proportion to the amount of this phase. These properties also depend upon structure.[26] The ferromagnetic condition is obtained by quenching the alloy from approximately 800° C. This treatment produces a body centered cubic structure with a face centered superlattice. Manganese is essential to the composition; copper may be replaced by silver; [26] aluminum by other trivalent or tetravalent elements—for example, tin.

Although possessing no properties of commercial utility, the alloys are extremely important in relation to magnetic theory. For those who may be particularly interested, a rather complete list [26] of recent papers on the subject is given in the references.

### *Summary of Magnetic Characteristics of Sheet Materials*

There are now available for magnetic purposes a large number of materials which are suitable for use over different ranges of magnetizing force and at different flux densities. In summarizing properties, the materials are grouped according to their adaptability at different magnetizing forces.

For low forces, where very high permeabilities are required, as in some signal apparatus, 78-permalloy is a suitable material. Molybdenum or chromium permalloy has properties similar to those of 78-permalloy but has higher initial permeability, is more readily heat-treated and is to be preferred for alternating current applications because of the higher resistivity. Hydrogen-treated iron at low magnetizing forces also has properties similar to those of the 78-permalloy.

For small magnetizing forces where a constant permeability and very low hysteresis loss is required, the heat-treated perminvars are available. The flux densities in cores of this type, if constant permeability and low hysteresis loss are retained, must be kept below 1000 gauss.

For apparatus using higher magnetizing forces, iron-nickel alloys in the range of 40 to 55 per cent nickel, and the silicon steels are most suitable. These two materials are widely used in electric transformers.

In the region of fields of 10 to 50 oersteds and higher, the 50–50 iron-cobalt and vanadium-modified alloy are attractive because of the high intrinsic induction of approximately 22,000 gauss. This superiority exists even in fields of over 1000 oersteds, but in the higher

ranges, the saturation value is approached by that of the less expensive material, magnetic iron. In connection with high flux density applications, the permalloys, the perminvars, and modified alloys of this character are of little interest since they are saturated at comparatively low flux densities.

The data given here are not sufficient to select a material for a specific application since many other detailed properties must be considered in connection with each individual problem. The materials which have been described, however, cover, in a general manner, the entire range of magnetic fields.

### *Magnetic Alloys in Dust Form*

For certain purposes where a substantially constant low permeability is desired, and in particular, for high frequency applications where eddy current losses are of consequence, it is desirable to produce the magnetic material in a fine powder, which is subsequently insulated and pressed into the desired core shape. One alloy prepared in this form is the permalloy [27] containing approximately 80 per cent of nickel. This is a material suited for use at audio frequencies, for example in the cores of loading coils.

The alloy may be prepared in powdered form by a number of methods. In one method, which is essentially metallurgical in nature, advantage is taken of the effect of small amounts of added elements.[27] It has been found that the addition of a few thousandths of a per cent of sulfur to the iron-nickel alloy containing approximately 80 per cent of nickel produces a structure which can be hot-rolled to a small section, but which when cold, is exceedingly brittle and can be pulverized to a fine dust. The presence of other elements in small amounts also affects the properties of the alloy. One element, manganese, has an effect opposite to that of sulfur, and if present in sufficiently large amounts, nullifies the action of sulfur by producing a tough and malleable casting.

The structural behavior of sulfur and manganese in permalloy is interesting in explaining the embrittling action. Sulfur exists in the structure as microscopic films of complex sulfides at the crystallite boundaries as illustrated in Fig. 13. These films are brittle, and when they extend over the greater portion of the crystallite surface, produce an interface of weakness permitting easy pulverization. The satisfactory hot-working properties can be explained in that at the high temperature, either the sulfide film is malleable, or dissolves in the iron-nickel solid solution. If manganese is added to an alloy containing sulfur, the sulfide constituent is blackened, loses its continuous

characteristics, and becomes agglomerated as is well known in the case of iron and of nickel. In this form it has a minimum embrittling effect since contact between metallic crystallites exists over the greater proportion of the interface. The structure is illustrated in Fig. 14. If sufficient manganese is present, alloys containing fairly high percentages of sulfur can not only be hot worked, but are malleable when cold.

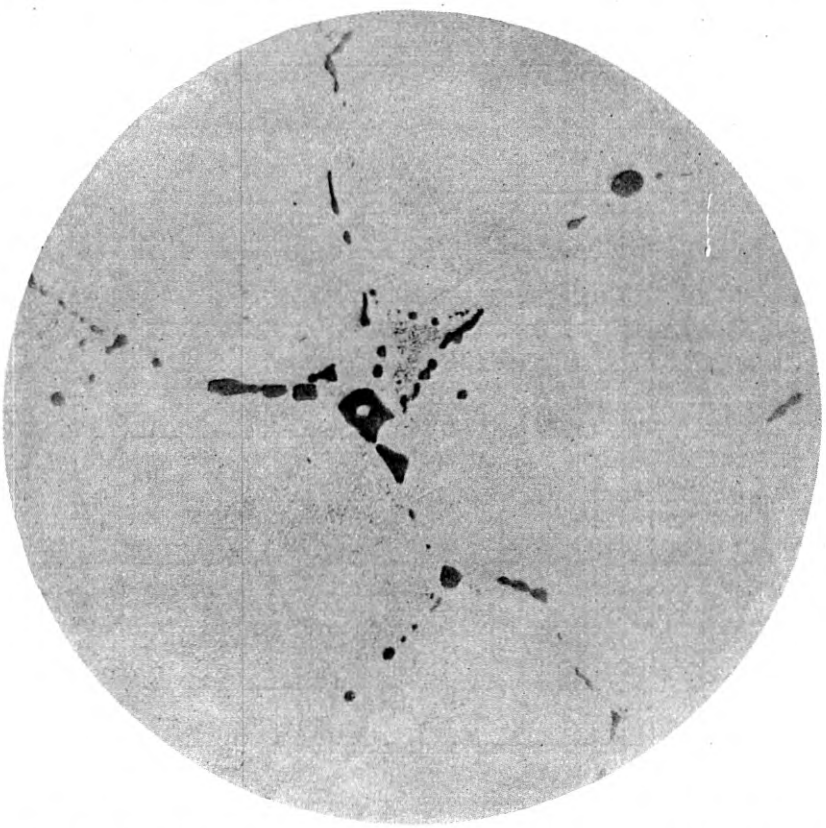

Fig. 13—Structure of an iron-nickel alloy containing 80 per cent nickel to which had been added 0.12 per cent sulphur and 1.05 per cent manganese. The sulfide constituent exists along the crystallite boundaries as grayish rounded inclusions. Mag. 2000 ×.

The production of magnetic dusts is unique as a metallurigcal process since the metallurgist is concerned with adding elements yielding a product which can be pulverized to a fine dust. Generally, the metallurgist exercises his greatest ingenuity to produce ductility and malleability in the end product.

## Hard Magnetic Materials (Permanent Magnets)

Permanent magnet materials, as pointed out in the introduction to this paper, are characterized by properties different from those of the soft magnetic alloys. With these materials, a high residual induction and high coercive force are required. Permanent magnets in use

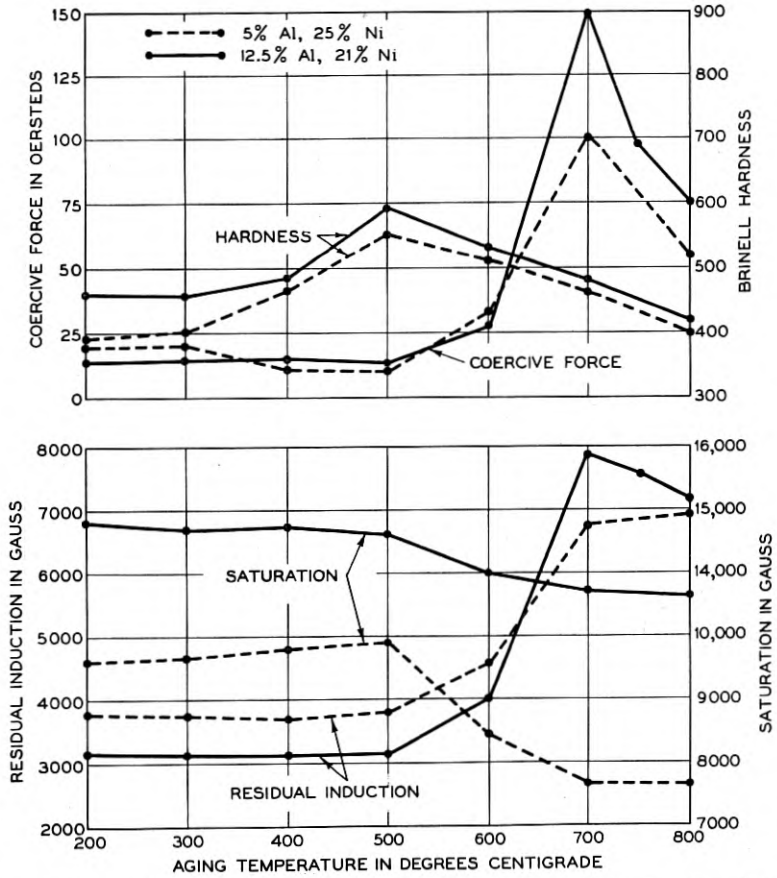


Fig. 14—Effect of aging treatment on typical iron-nickel-aluminum magnet alloys—Köster[31].

commercially at this time are steels. High intrinsic induction is assured by a high proportion of iron in the composition; the magnetic hardness is assured by the addition of carbon, which, when the steel is suitably heat-treated, precipitates as a finely dispersed carbide throughout the matrix. This finely dispersed precipitate keys the

structure and furnishes resistance to change in magnetic condition, a property which is measured as coercive force.

*Permanent Magnet Steels*

Many types of steel have been developed for magnet applications.[28] The most important commercial ones are carbon-manganese, 1 per cent and higher chrome steels of which 3.5 chrome is typical, tungsten steel, and 35 per cent cobalt steel. Typical magnetic characteristics and required heat treatments are shown in Table I.

TABLE I

THE PROPERTIES OF PERMANENT MAGNET STEELS

| Type of Steel | Typical Composition Per Cent | Heat Treatment | Magnetic Characteristics | | |
|---|---|---|---|---|---|
| | | | $H_c$ | $B_r$ | $H_c \times B_r \times 10^{-3}$ |
| Carbon-Manganese Steel..... | C—0.60<br>Mn—0.80<br>Si—0.20 | Quenched 800° C. in water | 40 | 9,000 | 360 |
| Low Chrome Steel | Cr—0.90<br>C—0.60 | Quenched 800° C. in oil | 50 | 9,500 | 475 |
| 3.5% Chrome Steel | Cr—3.5<br>C—0.90 | Quenched 830° C. in oil | 68 | 9,500 | 645 |
| Tungsten Steel.... | W—5.0<br>C—0.7 | Quenched 845° C. in water | 60 | 10,500 | 630 |
| 35% Cobalt Steel. | Co—35<br>W—7<br>Cr—4<br>C—0.75 | Quenched 940° C. in oil | 220 | 9,500 | 2090 |

The heat treatment for the permanent magnet steels consists of a quench from a high temperature at which the carbon is in solution. During the period of cooling in the quench, the supersaturated solution precipitates the carbides. This is designated "quench hardening." The quenching treatment varies with the type of steel, but in general, quenching in either oil or water is used. In this connection, it is important that the austenite transformation occur during the quench, since austenite is non-ferromagnetic and its presence will result in low values of residual induction.

The carbon-manganese and 1 per cent chrome steels are the least expensive and enjoy a large use in low cost apparatus where space requirements are liberal. Typical values for coercive force and residual induction for these materials are 40–50 oersteds and 9000 to 9500

gauss. The 1 per cent chrome steel has the advantage over the carbon-manganese steel in that the desirable characteristics are produced by oil quenching while the carbon-manganese steel is ordinarily water-hardened. Water-hardening frequently results in cracked magnets and consequently a higher proportion of rejections.

Of slightly higher quality are the 3.5 per cent chrome and the 5 per cent tungsten steels.* Typical values of coercive force are from 60 to 70 oersteds, and of residual induction from 9500 to 10,500 gauss. In general, tungsten steel in the hardened condition has a higher residual induction than the other magnet steels. Tungsten steel requires water-quenching, while 3.5 chrome steel is oil-quenched. In addition, chrome steel is a somewhat lower cost material. For these reasons, in recent years it has been substituted to some extent in applications where tungsten steel was formerly used.

The best permanent magnet steel in commercial use is the 35 per cent cobalt steel,[29] a complex alloy which contains, in addition to the cobalt, tungsten, chromium and carbon. Typical values of coercive force and residual induction for this material are 220 oersteds and 9500 gauss. Although this material is decidedly superior in properties to the other magnet steels, because of high cost, its use is limited to applications where space curtailment and apparatus requirements eliminate the cheaper steels.

*New Developments in Permanent Magnet Alloys*

Within the last five years, there have been a number of publications [30, 31] describing new materials which have properties of interest to engineers using permanent magnets. These materials are alloys with no intentional carbon additions, and, hence, are a radical departure in this field. The new magnet alloys solidify as alpha-solid solutions which by suitable heat treatment at a lower temperature decompose precipitating second phases. Contrary to the case with the iron-carbon alloys, the alpha-solid solution undergoes no phase change with decreasing temperature. Consequently the alloys have a coarse grain while the hardened magnet steels have a fine grained structure due to the intermediate phase change.

The permanent magnet qualities, however, result from the same type of metallurgical reaction that occurs in the carbon steels, that is, a precipitation of a second phase which is dispersed throughout the alpha-solid solution. The useful properties are secured by the usual precipitation hardening treatments; either a quench from a high

* A commendable detailed discussion of tungsten magnet steels is given by Gregg in the recent book, Alloys of Iron and Tungsten, p. 212, McGraw Hill, 1934.

temperature during which precipitation occurs, or a quench followed by an aging treatment at a lower temperature to produce the precipitation. Examples of alloys of this type are the iron-nickel-aluminum alloys, which have been described by Mishima,[30] and the iron-cobalt-molybdenum and iron-cobalt-tungsten alloys, which have been described by Seljesater and Rogers [31] and Köster.[31]

### *Iron-Nickel-Aluminum Alloys*

A representative composition of the alloys described by Mishima [30] consists of 65 per cent iron, 25 per cent nickel and 10 per cent aluminum. The composition may be further modified by the addition of manganese, vanadium, cobalt, chromium, tungsten, molybdenum, or copper.

For the simple ternary alloy in the cast condition a coercive force of 240 oersteds and a residual induction of 9600 gauss have been reported by Mishima.[30] By slight modifications in compositions, coercive forces of over 500 oersteds in combination with residual inductions of approximately 9500 gauss are reported. Values for three Mishima alloys, presumably of different composition, have been reported recently by Steinhaus and Kussman [30] and are given below.

| Type of Mishima Alloy | Coercive Force Oersteds | Residual Induction Gauss |
|---|---|---|
| MK1 | 660 | 7,600 |
| MK3 | 440 | 9,800 |
| MK5 | 130 | 10,800 |

Köster [30] has investigated the ternary equilibrium conditions for the iron-nickel-aluminum alloys. In the range of compositions of interest for magnet purposes, a surface of solubility varying with the temperature exists. It would be expected, therefore, that these alloys would be amenable to age hardening treatment. This has actually been demonstrated by Köster for the alloys of iron-nickel-aluminum. The curves shown in Fig. 14 are reproduced from his published data and will be recognized as demonstrating typical age hardening phenomena. The optimum aging temperature appears to be 700° C.

The fact that Mishima obtained high permanent magnet quality in specimens in the cast condition can be explained in that the precipitation of the second phase occurs during the simple cooling of the casting. It would be expected, therefore, that the magnet properties obtained would depend upon the casting dimensions and the rate of cooling. If this is true, it might be desirable in some magnet structures to subsequently heat treat the material to obtain uniform and reproducible results.

The fact that precipitation occurs is indicated more completely by examination of the photomicrographs in Figs. 15 and 16 of a typical

iron-nickel-aluminum alloy after quenching, and after quenching followed by aging at 1000° C. for 24 hours. There is no visual evidence of the second phase in the quenched specimen although some precipitation of submicroscopic particles undoubtedly has occurred. In the aged specimen, in Fig. 16, the second phase appears unmistakably.

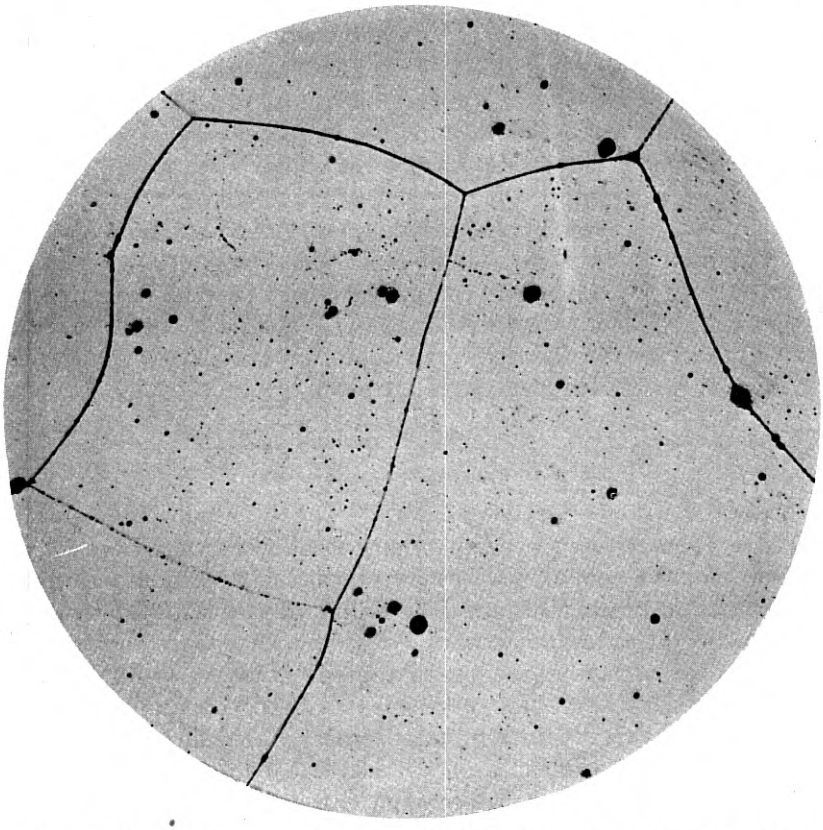

Fig. 15—Typical structure of an iron-nickel-aluminum alloy, containing 25 per cent nickel and 10 per cent aluminum, after quenching from 1200° C. in oil. In this condition the alloy has its optimum permanent magnet quality. Mag. 125 ×.

*Iron-Cobalt-Molybdenum and Iron-Cobalt-Tungsten Alloys*

The age hardening characteristics of alloys in these systems were established some years ago by Sykes, who developed the alloys for tools. Later, Seljesater and Rogers [31] reported that these alloys possess permanent magnet characteristics. A detailed report on the magnet properties has been made by Köster.[31] Of the two systems, it appears

that appreciably higher coercive forces are available in the one consisting of iron, cobalt and molybdenum. In Fig. 17 are shown the combinations of residual induction and coercive force obtained by Köster in the latter system with variations in molybdenum and cobalt content. Coercive forces of over 200 oersteds, associated with residual inductions of approximately 10,000 gauss, are given.

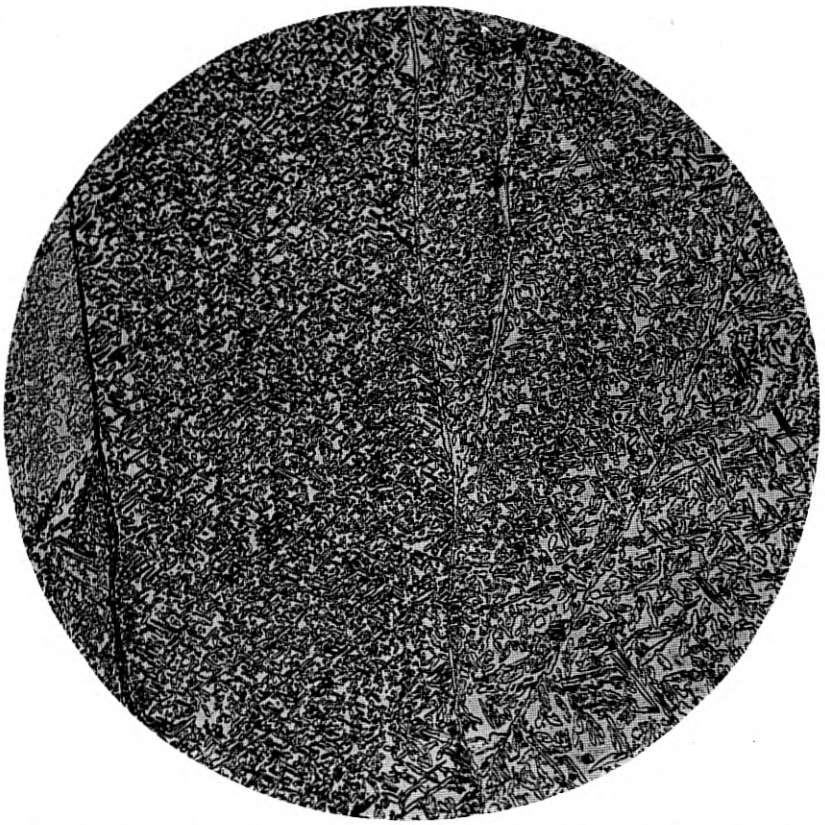

Fig. 16—Structure of the same alloy shown in Fig. 15 after a subsequent aging treatment of 24 hours at 1000° C. This treatment has resulted in the precipitation of a large amount of the second phase. Mag. 125 ×.

The desirable properties of this class of materials are developed by a double heat treatment which consists of a quench from a high temperature followed by aging at a lower temperature. The alloys are precipitation hardening in the same manner as the non-ferrous copper-beryllium alloys or lead-calcium cable sheath alloys. The heat treatment is illustrated in Fig. 18, which shows the effect of aging on the

properties of an alloy containing 15 per cent cobalt and 18 per cent tungsten reproduced from Köster's published data. The alloy was initially quenched from a temperature of 1300° C., and subsequently

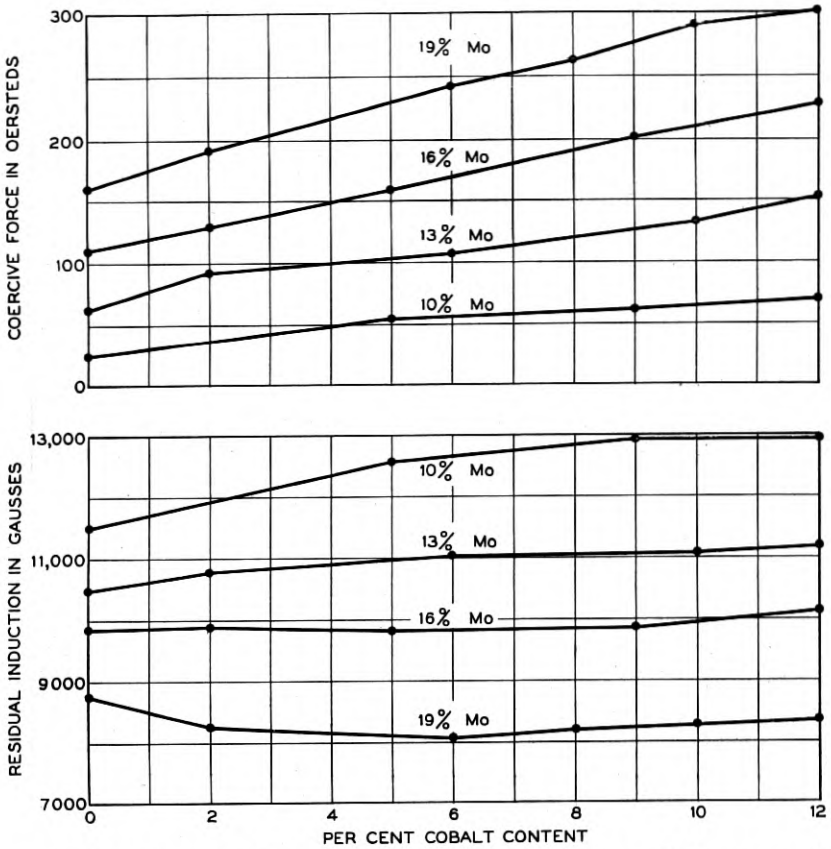


Fig. 17—Effect of different combinations of cobalt and molybdenum on the coercive force and residual induction of iron-cobalt-molybdenum alloys in the heat-treated condition—Köster[31].

aged at temperatures from 300 to 800° C. The best results are obtained by aging at temperatures between 700° C. and 800° C.

The precipitation of the second phase in a typical iron-cobalt-molybdenum alloy is illustrated by a comparison of Figs. 19 and 20. The section in Fig. 19, which is for the alloy after quenching from 1300° C., shows a typical solid solution structure with a small number of rounded inclusions resulting either from the deoxidation of the melt or from incomplete solution at the high temperature. In Fig. 20, a section

of the same alloy after quenching and subsequent aging, the precipitate is visible as rounded particles within the grains and as elongated, needle-like structures within and along grain boundaries.

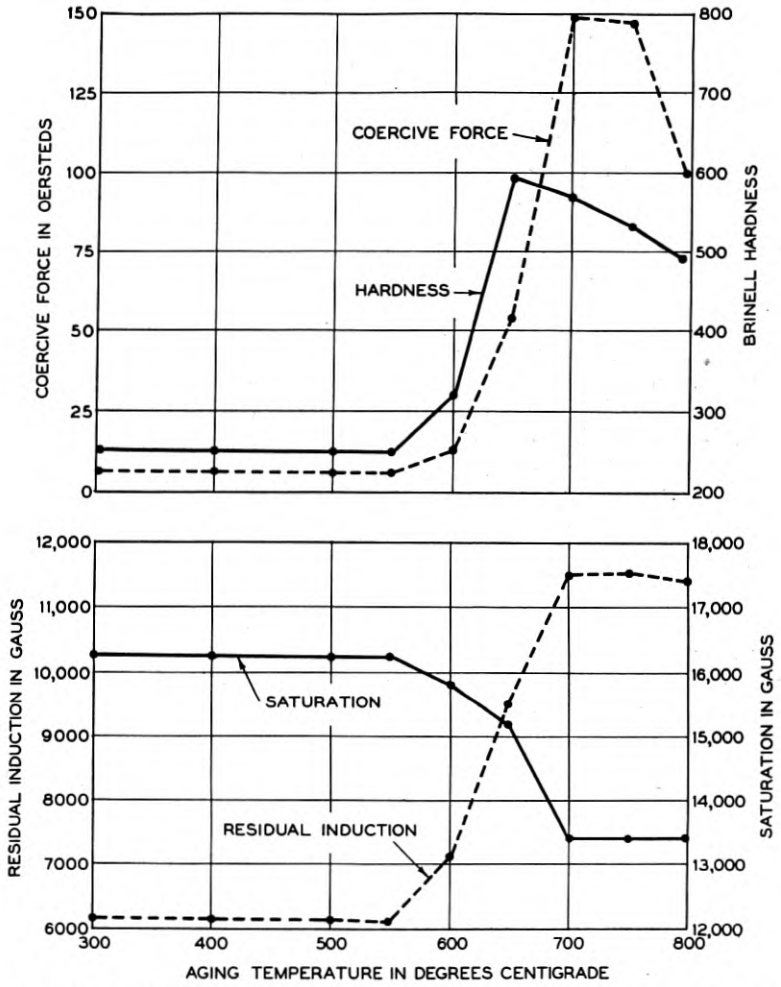


Fig. 18—Effect of aging treatment showing typical age-hardening phenomena for an iron-cobalt-tungsten magnet alloy—Köster[31].

Köster [31] reports that these alloys can be rolled or otherwise shaped at sufficiently high temperatures and can be readily stamped at red heat. The alloys machine readily in the quenched condition but on account of the coarse grained structure and consequent brittleness they must be machined without shock.

*Oxide Magnets*

In this class of materials are included the metallic oxide magnets recently described in the literature by Kato and Tokei.[32] The metallic oxide magnets open a new field of permanent magnet materials.

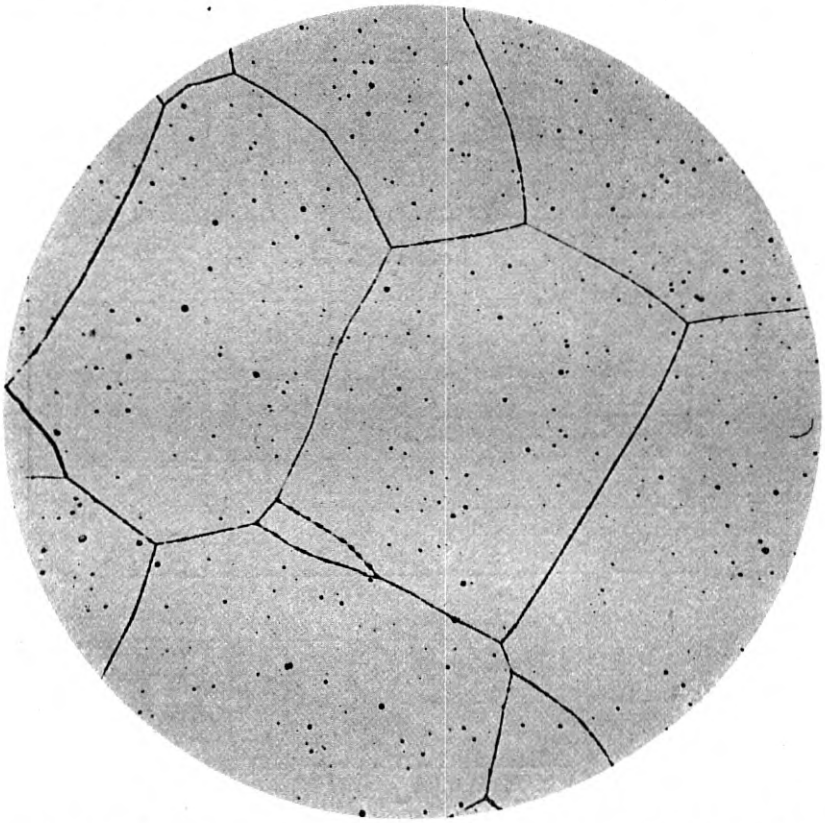

Fig. 19—Typical structure of an age-hardening iron-cobalt-molybdenum alloy, containing 12 per cent cobalt and 16 per cent molybdenum, after quenching from 1300° C. Mag. 125 ×.

These magnets are composed, according to the Japanese authors, in one instance of a solid solution of cobalt ferrite ($CoFe_2O_4$) in magnetic iron oxide ($FeFe_2O_4$). Although the method of manufacture and exact compositions are not completely disclosed, it appears that the preparation involves the powdering of the metallic oxides, compressing in a suitable die to the desired shape, and subsequent heating.

Cobalt ferrite magnets, it is reported, are not easily magnetized at room temperature; but if the temperature is raised, for example, to

300° C. the magnetization is readily accomplished. In this procedure the magnet is cooled in the magnetizing field. If the magnet has once been magnetized at the high temperature, succeeding magnetizations can be carried out at room temperature.

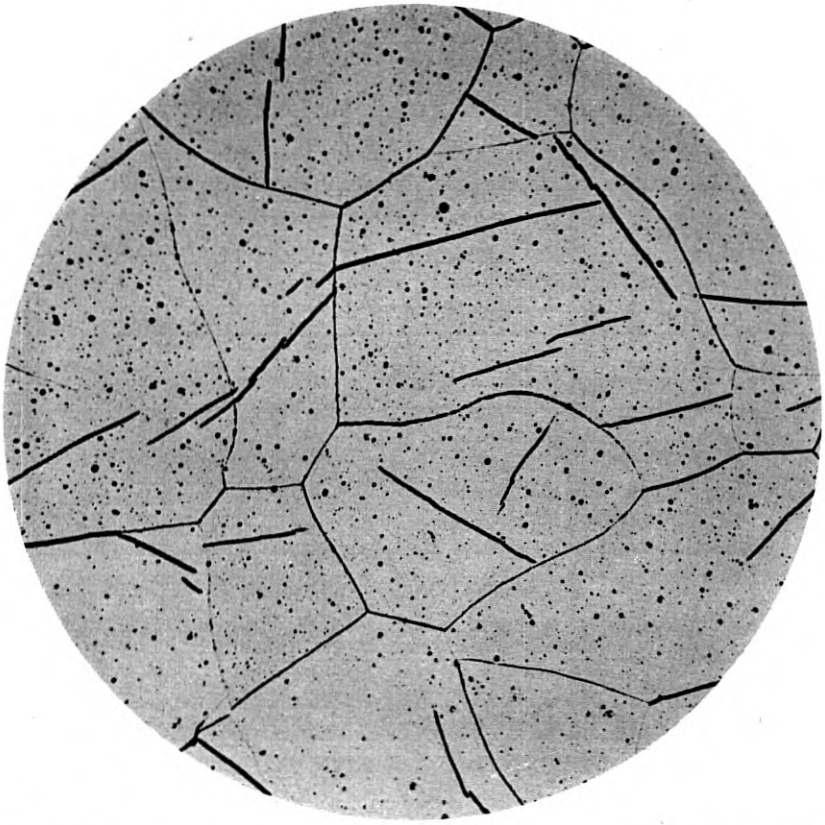

Fig. 20—The same alloy as shown in Fig. 19 after a subsequent aging treatment of 5 hours at 725° C. Mag. 125 ×.

The magnetic characteristics of the material are illustrated in Fig. 21, in which a hysteresis loop for a metallic oxide magnet is compared with one for carbon steel. The coercive force of the new material is approximately 600 oersteds and the residual intensity of magnetization, 320 units, corresponding to a residual induction of approximately 4000 gauss.

The high coercive force of the oxide magnet makes possible the more exact location of poles and permits the utilization of the material in

short magnets. Because of the lower density of the oxide compared with that of iron, the oxide magnets have approximately the same induction flux as metallic magnets on a weight basis.

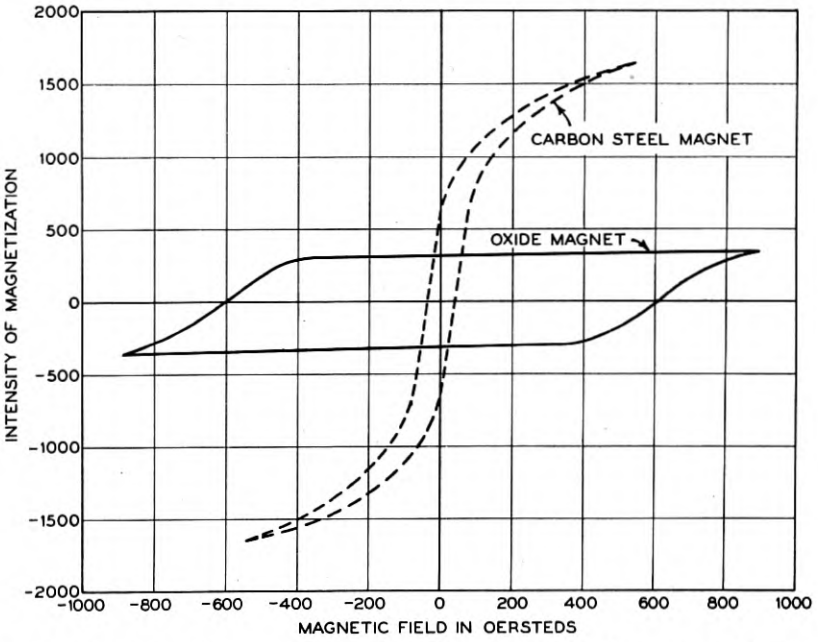


Fig. 21—Comparison of some magnetic characteristics of an oxide magnet with those of carbon steel from Kato and Tokei[32].

The description of the materials given here has been obtained from the Japanese publication.[32] So far as is known, no investigation of the preparation or of the properties of the material has been reported in this country.

*Summary of New Permanent Magnet Materials*

The properties of the new permanent magnet materials discussed are summarized in Table II.

TABLE II

THE PROPERTIES OF SOME NEW PERMANENT MAGNET MATERIALS

| Materials | $H_c$-Oersteds | $B_r$-Gauss | $H_c \times B_r \times 10^{-3}$ |
|---|---|---|---|
| Iron-Cobalt-Molybdenum Alloys.... | 50–300 | 12,000–8,500 | 600–2,550 |
| Iron-Cobalt-Tungsten Alloys..... | 150 | 11,500 | 1,720 |
| Iron-Nickel-Aluminum Alloys..... | 130–660 | 10,800–7,600 | 1,400–5,000 |
| Oxide Magnets..... | 600 | 4,000 | 2,400 |

An important consideration in connection with the new materials is the broadening of the field of possible new magnet alloys. Previously it had been generally considered that the field of useful permanent magnet materials was confined to the plain and alloy steels. The new metallic materials are systems which are precipitation hardening, and suggest possibilities for developing useful alloys in other similar systems of iron.

REFERENCES


1. Scott, K. L. "Magnet Steels and Permanent Magnets," *Elec. Engg.*, **51**, 320 (1932).
   Evershed, S. "Permanent Magnets in Theory and Practice," Jour. I. E. E., **63**, 725 (1925).
2. Messkin-Kussmann. "Die Ferromagnetischen Legierungen," p. 308, Julius Springer, Berlin (1932).
3. Scott, H. "Solubility of Carbon in Alpha Iron," *Chem. and Met. Engg.*, **27**, 1156 (1922).
   Whiteley, J. H. "Solution of Cementite, in $\alpha$-Iron and Its Precipitation," *Jour. Iron and Steel Inst.*, **116**, 293 (1927).
   Bates, A. A. "Aging in Low Carbon Steels," *Trans. Amer. Soc. Steel Treat.*, **19**, 449 (1932).
4. Tritton, F. S. and Henson, D. "Ferrous Alloys Research—Pt. II—Iron and Oxygen," *Jour. Iron and Steel Inst.*, **110**, 90 (1924).
   Jordan, L. and Eckman, J. R. "Gases in Metals II—Determination of Oxygen and Hydrogen in Metals by Fusion in Vacuum," *Sci. Papers, Bu. Stds.*, Paper No. 514, **20**, 445 (1925).
5. Epstein, S. "Observations on the Iron-Nitrogen System," *Trans. Amer. Soc. Steel Treat.*, **16**, 19 (Nitriding Symposium), (1929).
6. Benedicks, C. and Löfquist, H. "Nonmetallic Inclusions in Iron and Steel," p. 13, Chapman and Hall, London.
7. Barrett, W. F., Brown, W. and Hadfield, R. A. "On the Electrical Conductivity and Magnetic Permeability of Various Alloys of Iron," Parts I and II, *Sci. Trans., Roy. Dublin Soc.*, Ser. 2, **7**, 67 (1900).
   Yensen, T. D. "Magnetic and other Properties of Iron Silicon Alloys Melted in Vacuo," *Illinois Univ. Eng. Expt. Stat. Bull.* No. 83 (1915).
8. Messkin-Kussmann. "Die Ferromagnetischen Legierungen," p. 312, Julius Springer, Berlin (1932).
9. Yensen, T. D. "Pure Iron and Allotropic Transformations," *Trans. Iron and Steel Div., A. I. M. E.*, p. 320 (1929).
10. Rohn, W. "Die Entwicklung der Heraeus—Vacuum Schmelze A-G. von 1923–1933," Heraeus Vacuum-Schmelze, 1923–1933, p. 1, Hanau, 1933.
11. Cioffi, P. P. "Hydrogenized Iron," *Physical Review*, **39**, 363 (1932).
    Cioffi, P. P. "New High Permeabilities in Hydrogen-Treated Iron," *Physical Review*, [2] **45**, 742 (1932).
12. Iron Carbonyl Patents: British No. 244,895; German No. 518,387; U. S. Nos. 1,783,744, 1,759,268, 1,725,619.
    Fieldner, A. C. and Jones, G. W. "Iron Carbonyls: Their Physical and Chemical Properties," *Amer. Gas Assoc. Monthly*, **6**, 439 (1924).
13. Buddenberg, O., Duftschmid, F., u. Schlecht, L. "Reines Carbonyleisen als hochpermeabler magnetischer Werkstoff," Heraeus Vacuum—Schmelze 1923–1933, p. 74, Hanau, 1933.
14. Keinath, G. "Hochmagnetische Legierungen aus Nickeleisen," *Archiv für technisches Messen*, **2**, T 173 (1932).
15. Pettit, I. C. "Magnetic Materials," *Bell Laboratories Record*, **12** (1934).
16. Yensen, T. D. "Magnetic Properties of the Fifty Per Cent Iron-Nickel Alloy," *Jour. Franklin Inst.*, **199**, 333 (1925).
17. Arnold, H. D. and Elmen, G. W. "Permalloy, an Alloy of Remarkable Magnetic Properties," *Jour. Franklin Inst.*, **195**, 621 (1923).
    Elmen, G. W. "Magnetic Alloys of Iron, Nickel, and Cobalt," *Jour. Franklin Inst.*, **207**, 583 (1929).

Elmen, G. W. "New Permalloys," *Bell Laboratories Record*, **10**, 2 (1931).
18. Dahl, O. "Zur Frage Unterkühlbarer Zustandsänderungen in Eisen-Nickel-legierungen," *Zeit. f. Metallkunde*, **24**, 107 (1932).
19. Haughton, J. L. and Payne, R. J. "Transformations in the Gold-Copper Alloys," *Jour. Inst. of Metals*, **46**, No. 2, 457 (1931).
20. Kussmann, A., Scharnow, B., u Steinhaus, W. "Über das Permalloy-Problem," Heraeus Vacuum—Schmelze 1923–1933, p. 310, Hanau, 1933.
Dahl, O. and Pfaffenberger, J. "Beitrag zur Kenntnis der Eisen—Nickel legierungen," *Zeit. f. Metallkunde*, **25**, 241 (1933).
21. Dahl, O., Pfaffenberger, J. and Sprung, H. "Neurartige magnetische Werkstoffe für pupinspulen (Materialien höchster magnetischer Stabilität; Ispoerme)," *Elektrische Nachrichten Technik.*, **10**, 317 (1933).
22. Ellis, W. C. "A Study of the Physical Properties of Electrolytic Cobalt and its Alloys with Iron," *Renss. Poly. Inst. Eng. and Sci. Series* No. 16, Troy, N. Y., 1927.
Kussmann, A., Scharnow, B., u. Schulze, A. "Physikalische Eigenschaften und Struktur des Zweistoffsystems Eisen-Kobalt.," *Zeit. f. technische Physik*, **13**, 449 (1932).
23. White, J. H. and Wahl, C. V. U. S. Patent No. 1,862,559.
24. Elmen, G. W. "Magnetic Properties of Perminvar," *Jour. Franklin Inst.*, **206**, 317 (1928).
Auwers, O., u Kühlewein, H. "Beitrage zum Stereomagnetismus II—Über das Perminvar-problem," *Ann. der Physik*, [5] **17**, 121 (1933).
25. Kelsall, G. A. "Permeability Changes in Ferromagnetic Materials Heat Treated in Magnetic Fields," *Physics*, **5**, 169, 1934.
Bozorth, R. M., Dillinger, J. F. and Kelsall, G. A. "Magnetic Materials of High Permeability Attained by Heat Treatment in a Magnetic Field," *Physical Review*, [2], **45**, 742 (1934).
McKeehan, L. W. and Cioffi, P. P. "Magnetostriction in Permalloy," *Physical Review*, [2] **28**, 146 (1926).
McKeehan, L. W. "Ferromagnetism in Metallic Crystals," *Metals Technology*, p. 1, August 1934.
Bozorth, R. M. "Theory of the Heat Treatment of Magnetic Materials," *Phys. Review*, **46**, 232 (1934).
26. Heusler, F. "Über den Zusammenhang der magnetischen und mechanischen Eigenschaften der gewalzten Heusler-Bronz (Mangan-Aluminium-Kupfer)," *Zeit. f. Physik.*, **10**, 403 (1922).
Persson, E. "Über den Bau der Heuslerschen Legierungen," *Zeit. f. Physik.*, **57**, 115 (1929).
Potter, H. H. "Some Magnetic Alloys and Their Properties," *Phil. Mag.*, [7] **12**, 255 (1931).
Bradley, A. K. and Rogers, J. W. "The Crystal Structure of the Heusler Alloys," *Proc. Roy. Soc. A.*, **144**, 340 (1934).
Heusler, O. "Kristallstruktur und Ferromagnetismus der Mangen-Aluminium-Kupferlegierungen," *Annalen der Physik*, [5] **19**, 155 (1934).
27. Shackelton, W. J. and Barber, I. G. "Compressed Powdered Permalloy," *Trans. Amer. Inst. Elec. Eng.*, **47**, 429 (1928).
White, J. H. U. S. Patent No. 1,739,052.
Beath, C. P. and Heincke, H. M. E. U. S. Patent No. 1,669,649.
Andrews, J. W. U. S. Patent No. 1,703,287.
28. Sanford, R. L. "Some Principles Governing the Choice and Utilization of Permanent Magnet Steels," *Sci. Papers, Bu. Stds.* Paper No. 567, **22**, 557 (1927).
Mathews, J. A. "Retained Austenite—A Contribution to the Metallurgy of Magnetism," *Trans. Am. Soc. Steel Treat.*, **8**, 565 (1925).
29. Honda, K. and Saito, S. "On K. S. Magnet Steel," *Physical Review*, [2] **16**, 495 (1920).
30. Mishima, T. "Magnetic Properties of Iron-Nickel-Aluminum Alloys," *Ohm* (Coversheet), July, 1932. *Iron Age* (Abstract), p. 346, Sept. 1932.
French Patent No. 731,361.
English Patents Nos. 392,656; 392,657; 392,658; 392,659; 392,660 and 392,661.
Köster, W. "Das System Eisen-Nickel-Aluminium," *Archiv. für das Eisen-hüttenwesen*, **7**, 257 (1933).

Steinhaus, W. u Kussmann, A. "Neuere Entwicklung der Dauermagnetstähle," *Phys. Zeitschrift*, **35**, 377 (1934).
"Nickel-Aluminium Magnetstahl," *Archiv. für technisches Messen*, **3**, T56 (1934).
31. Seljesater, K. S. and Rogers, B. A. "Magnetic and Mechanical Hardness of Dispersion Hardened Iron Alloys," *Trans. Amer. Soc. Steel Treat.*, **19**, 553 (1932).
Köster, W. "Über die Beziehungen der magnetischen Eigenschaften, insbesondere der Koerzitivkraft zum Gefügeaufbau der Legierungen und die Entwicklung neuartiger Magnetlegierungen," *Zeit. f. Elektrochemie u. Angewandte Physikalische Chemie*, **38**, 549 (1932).
Köster, W. "Dauermagnetiwerkstoffe auf der Grundlage der Ausscheidungshärtung," *Stahl u. Eisen*, **53**, 849, 1933.
32. Kato, Y. and Tokei, I. "Permanent Oxide Magnet and Its Characteristics," *Jour. Inst. of Elec. Eng. of Japan*, **53**, 408 (1933).

# Theory of Multi-Electrode Vacuum Tubes *

By H. A. PIDGEON

Physical principles underlying the characteristics and performance of multi-electrode vacuum tubes are presented in simple form in this paper. Presentation of the subject is based, as far as possible, on the well-known theory of the three-electrode tube. It is shown that the definitions of electrical tube parameters applicable to triodes are, with certain modifications in their interpretation, also applicable to tubes having more than three electrodes.

The characteristics of the screen-grid tetrode are discussed in detail since they are typical of those found in a number of multi-electrode structures. Except for the effects produced by the emission of secondary electrons from the plate and screen, it is shown that the characteristics of the screen-grid tetrode are in general accord with those to be expected from the application of simple theory. The presence of the electrostatic screen in such structures inherently results in high values of the plate resistance and amplification factor, but the transconductance remains normal and has about the same value as in comparable triodes.

One of the necessary modifications in the screen-grid tetrode to produce a satisfactory output power tube is some means of removing the fold in the plate current-plate voltage characteristics, which limits the permissible plate voltage swings. This is accomplished in the power pentode by the addition of a suppressor grid between the plate and screen grid. The efficiency of power pentodes, and of some other tubes having positive grids, is higher than that usually found in triodes. The reason for this is discussed and also certain peculiarities in the harmonic output of pentodes.

The arrangement of electrodes in the space-charge-grid pentode corresponds to that of the screen-grid tetrode with an additional grid inserted between the cathode and control grid. This space-charge grid, which is maintained at a positive potential of 10 to 20 volts with respect to the cathode, reduces the effects of space charge near the cathode surface. This results in extraordinarily high values of transconductance and, consequently, in high amplification. Practically, such tubes are limited to use as voltage amplifiers, since operation over the wide range necessary for large output power results in prohibitive distortion.

In the co-planar-grid tetrode the lateral wires of the positive grid are arranged in the same planes as those of the control grid. This results in comparatively low plate resistance while retaining the advantages of a positive grid. The plate efficiency is comparable with that in the power pentode but the available amplification is lower.

INTRODUCTION of the three-electrode vacuum tube into the field of communications and in other applications represented such a tremendous advance over the possibilities of any other known device that, despite some of its rather obvious limitations, it proved entirely adequate for the service required until comparatively recent years. However, with increasing demands made by service requirements for larger power output at higher efficiency, reduced distortion, higher

* Published in November 1934 issue of *Electrical Engineering*. Scheduled for presentation at Winter Convention of A. I. E. E., New York City, January 22–25, 1935.

gain, amplification at higher frequencies and greater frequency discrimination, it eventually became necessary to investigate the possibilities of making changes in vacuum tubes enabling them to meet these requirements more satisfactorily. Measures taken to meet this situation have included improvements in the three-electrode tube that reduce the effects of some of its limitations, and the development of vacuum tubes having more than three electrodes.

The purpose of this paper is to present, in simple form, the physical principles underlying the characteristics and performance of multi-electrode vacuum tubes. For present purposes, such tubes may be defined as those having more than the three electrodes of the conventional triode. The procedure will be to show that the definitions of electrical tube parameters applicable to triodes are, with certain modifications in their interpretation, also applicable to tubes having more than three electrodes; and, utilizing the theory of the triode, to analyze the characteristics of a few typical multi-electrode structures that illustrate the types of characteristics found in many such tubes. No attempt is made to present new material in the paper or to discuss in detail the many different types of multi-electrode tubes now in use. The author has attempted to present the subject from the viewpoint of those readers who have a satisfactory understanding of the physical principles, characteristics, and operation of the triode, but who do not have a similarly clear analysis available for the more complex structures.

Multi-electrode tubes may be divided conveniently into two classes. In the first class are those the purpose of which is to perform some function that cannot be performed readily by a triode, or which perform some function better by reason of the elimination or reduction of some limitation in triodes. The second class includes those structures in which additional electrodes are introduced to permit them to perform simultaneously more than one function, or to permit them to function in two or more ways, depending on the voltages applied to the various electrodes and on the manner of their operation. This paper will deal exclusively with typical structures of the first class.

## Fundamental Definitions and Tube Equations

Regardless of the type of multi-electrode tube considered, the space current to any electrode may be expressed as some function of the voltages applied to the various electrodes. However, in the operation of any multi-electrode device, or any section of such a device performing a single function, one usually is concerned with variations in the voltages and corresponding currents of only two of the electrodes, the other electrodes being maintained at fixed potentials. One of these

two electrodes, which usually is maintained at a negative operating potential, is connected to the input circuit and acts as a control electrode. The second of these two electrodes, which is maintained at a positive potential with respect to the cathode, acts as an anode or collector of electrons and is connected in the output circuit. Just as in the case of the triode, then, a study of the characteristics of multi-electrode tubes is concerned with variations in the current collected by the anode with variations in the potential applied to the control grid.

This anode or plate current may be expressed as a function of the various electrode voltages by the following equation:

$$I_p = f(E_p, E_{g_1}, E_{g_2}, E_{g_3}, \text{etc.}), \tag{1}$$

in which $E_p$ is the operating voltage applied to the output anode or plate, and $E_{g_1}$, $E_{g_2}$, $E_{g_3}$, etc., are the operating voltages applied to the various grids numbered outward from the cathode. The variation in the anode current, neglecting second and higher order terms, is given by

$$dI_p = \frac{\partial I_p}{\partial E_p} dE_p + \frac{\partial I_p}{\partial E_{g_1}} dE_{g_1} + \frac{\partial I_p}{\partial E_{g_2}} dE_{g_2} + \frac{\partial I_p}{\partial E_{g_3}} dE_{g_3} + \text{etc.} \tag{2}$$

The partial differential coefficients in equation 2 have the physical dimensions of conductances and, if these conductances be designated by the letter $S$ with appropriate subscripts, the equation may be written

$$dI_p = S_{pp} \cdot dE_p + S_{p1} \cdot dE_{g_1} + S_{p2} \cdot dE_{g_2} + S_{p3} \cdot dE_{g_3} + \text{etc.} \tag{3}$$

The plate or output anode conductance of a multi-electrode tube is defined in the same manner as for the three-electrode tube. It is the rate of change of plate current with plate voltage, that is, it is the slope of the plate current-plate voltage characteristic at the selected operating point, the potentials of all the other electrodes remaining constant. Under this condition, from equations 2 and 3

$$\text{Plate conductance} = \frac{\partial I_p}{\partial E_p} = S_{pp} = S_p. \tag{4}$$

Obviously, the plate resistance also must be defined in the same manner as in the triode, that is

$$\text{Plate resistance} = R_p = \frac{1}{\frac{\partial I_p}{\partial E_p}} = \frac{1}{S_p}. \tag{5}$$

In a similar manner, the transconductance from the control grid to

the output anode or plate of a multi-electrode tube is defined, as it is in the triode, by the rate of change of plate current with variation of the control-grid voltage; that is, it is the slope of the plate current-grid voltage characteristic at the given operating point, the potentials of all electrodes other than the control grid remaining constant.

In conventional screen-grid tetrodes and pentodes, the grid next to the cathode is the control grid. Consequently, for such structures, the transconductance is defined from equations 2 and 3 by

$$\text{Transconductance} = \frac{\partial I_p}{\partial E_{g_1}} = S_{p1}. \tag{6}$$

In space-charge-grid tetrodes and pentodes, the grid next to the cathode is maintained at a fixed positive potential and the second grid acts as the control grid. Consequently, in these and similar structures

$$\text{Transconductance} = \frac{\partial I_p}{\partial E_{g_2}} = S_{p2}. \tag{7}$$

Similarly, considering the control grid (assumed to be the first grid) and the output electrode of a multi-electrode tube, the amplification factor is defined, as it is in the triode, by the ratio of the transconductance to the plate conductance. It is expressed by

$$\text{Amplification factor} = \mu_{pg_1} = \frac{\dfrac{\partial I_p}{\partial E_{g_1}}}{\dfrac{\partial I_p}{\partial E_p}}. \tag{8}$$

Or, assuming that $E_{g_1}$ and $E_p$ are varied in such a manner that $I_p$ remains constant, the amplification factor is expressed in the usual form by

$$\mu_{pg_1} = -\left.\frac{dE_p}{dE_{g_1}}\right]_{I_p = \text{constant}}. \tag{9}$$

Combining equations 5 and 6 with equation 8 gives

$$\text{Transconductance } S_{p1} = \frac{\mu_{pg_1}}{R_p} \tag{10}$$

just as in the case of the triode. Exactly similar equations apply if $g_2$ is used as the control grid.

Obviously, the currents to the other electrodes in a multi-electrode tube may be expressed by functions of the electrode voltages, similar to equations 1 and 2. The various differential coefficients of these equa-

tions define transconductances, electrode resistances, and amplification (or reflex) factors analogous to those just given. Since these quantities are not used in this paper, they will not be given further consideration here. The voltage applied to the control grid will be designated by $E_g$, regardless of the grid employed for the purpose; and the transconductance (or mutual conductance) and the amplification factor, applying to the control grid and plate, will be designated by $S_m$ and $\mu$, respectively.

If a load resistance, $R$, is inserted in the plate circuit of a multielectrode tube, and if the potentials of all of the elements other than the control grid and plate are maintained constant, equation 3 reduces to

$$dI_p = S_p \cdot dE_p + S_m \cdot dE_g = \frac{1}{R_p} \cdot dE_p + \frac{\mu}{R_p} \cdot dE_g. \tag{11}$$

In this case, the only independent variable is $E_g$, and $E_p$ varies by reason of the changing potential drop across the external load resistance, $R$, due to variations in the plate current, $I_p$, produced by the varying grid potential. Consequently

$$dE_p = - dI_p \cdot R. \tag{12}$$

Substituting equation 12 in equation 11 and reducing,

$$dI_p = \frac{\mu}{R_p + R} \cdot dE_g. \tag{13}$$

For vacuum tubes having curvilinear characteristics, equation 13 applies rigorously, of course, only to infinitesimal variations in $I_p$ and $E_g$. However, as in the case of the triode, the output from multielectrode tubes may be expressed by a power series in terms of finite voltage variations applied to the elements, the coefficients in the series being functions of the static characteristics. If these finite variations in $I_p$ and $E_g$ are designated by $i_p$ and $e_g$, respectively, the output current is expressed to the first order by

$$i_p = \frac{\mu e_g}{R_p + R}, \tag{14}$$

which is identical with the equation expressing the output current from a triode.

Letting $e_p$ represent the variable voltage across the load resistance, $R$, the voltage amplification is given by

$$A_v = \frac{e_p}{e_g} = \frac{i_p \cdot R}{e_g} = \frac{R}{R_p + R} \cdot \mu. \tag{15}$$

It may also be written in the following form which will be found useful later:

$$A_v = \frac{R_p \cdot R}{R_p + R} \cdot \frac{\mu}{R_p} = \frac{R_p \cdot R}{R_p + R} \cdot S_m. \tag{16}$$

It should be emphasized, perhaps, that the electrical parameters of multi-electrode tubes and the output current, as defined by the foregoing equations, are subject to the condition that the voltages applied to all of the electrodes other than the plate and control grid are maintained constant. The satisfactory operation of multi-electrode tubes in circuits also usually requires that this condition be fulfilled. It requires that the impedance to alternating current components in each of these circuit branches be very low. This is accomplished in practice by connecting these electrodes to ground, so far as alternating currents are concerned, through reasonably large capacitances.

From the foregoing analysis it is apparent that, with proper interpretation, the definitions of plate resistance, transconductance, and amplification factor applicable to triodes are also applicable to multi-electrode tubes; in addition, the same expressions for output current and voltage amplification are applicable. This follows from the fact that these quantities are expressed in terms of the differential coefficients of the static characteristics, that is, they depend only upon the slopes of these characteristics at the given operating voltages and not upon their form. However, as will appear later, the difference in the shape of the static characteristics of multi-electrode tubes from those of triodes is very important in determining great differences not only in the magnitude of the electrical parameters, but also in the character and amount of distortion resulting when the tubes are operated under conditions such that large portions of the characteristics are traversed.

In multi-electrode tubes, as well as in triodes, the total space current drawn from the cathode is determined by the extent to which the resultant field, due to the electrodes, overcomes the opposing field produced by space charge. While space charge extends throughout the interelectrode space, it is relatively so much more dense in regions of very low electron velocity that, as a first approximation, its effect usually may be neglected in other regions. Except in space-charge-grid tubes and a few other special tubes, the only important space-charge region is confined to a relatively thin sheath near the cathode surface. Consequently, in such structures the total space current is determined largely by the extent to which the resultant positive field due to the electrodes neutralizes the negative field near the cathode

surface produced by space charge. An appreciation of this fact is essential to a clear understanding of the characteristics of multi-electrode tubes.

## Screen-Grid Tetrodes

Utilizing the simple theory of triodes, which has been shown to be applicable to multi-electrode tubes also, the characteristics of a multi-electrode tube will be analyzed next. For this purpose the screen-grid tube is chosen, since it admirably illustrates the type of characteristics found in several types of multi-electrode tubes.

The objective in this case is to reduce the direct capacitance between the plate and control grid through which energy is fed back from the plate to the input circuit. This is accomplished by inserting an

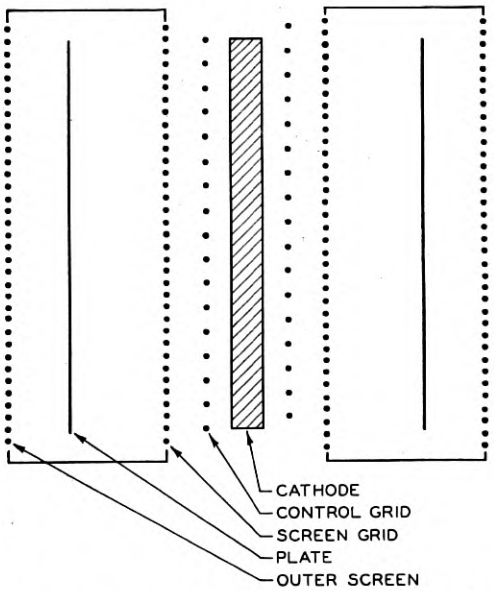


Fig. 1—Schematic diagram showing the arrangement of electrodes in a screen-grid tetrode.

electrostatic shield between the plate and control grid of what otherwise would be a three-electrode tube. The condition that such a screen must allow an electron stream to flow through it to the plate, is satisfied by making it of fine mesh or in the form of a finely wound grid structure. To be effective, it must be maintained at some constant potential with respect to ground.

The arrangement of the electrodes in such a tube is shown diagrammatically in Fig. 1. The screen structure outside the plate is added for

the purpose of completing the electrostatic isolation of the plate and its leads from the grid, thus reducing the capacitance between these two elements to the lowest practicable value. This outer screen is of no further concern, since it has no effect on the static characteristics of the tube. The usual characteristics of a typical screen-grid tetrode are shown in Figs. 2, 3, and 4. In this case, as is usual in screen-grid tubes, the shielding is very complete, reducing the direct capacitance between the plate and control grid to a few thousandths of what its value would be in the absence of the screen.

First, the characteristics of Fig. 2 will be considered. Since the direct capacitance between the plate and the control grid, $g_1$, is extremely small, the electric field in the immediate vicinity of the latter, produced by any potential on the plate, also must be extremely small. The cathode is electrically even more remote from the plate than the control grid, because it not only is shielded from the former by the screen grid, but also has some additional shielding from the control grid. Consequently, the field produced by the plate at points between the cathode and control grid is smaller even than the field produced at the control grid and, hence, is negligibly small. Since, as previously discussed, the total space current is determined almost wholly by the field very near the cathode surface, the plate in this case can have practically no effect on the total space current drawn from the cathode. This is shown by the curves giving the total space current, $I_t$, in Fig. 2. These curves are seen to be so flat as to be almost entirely independent of variations in the plate voltage.

The plate, then, in a screen-grid tube plays an essentially passive rôle which is to collect those electrons that succeed in passing through the screen. The remainder of the space current is collected by the screen, the sum of the plate and screen currents remaining nearly constant with changes in plate voltage.

There is nothing in the theory of the triode by which to determine the ratio in which space current divides between two or more positive electrodes in a multi-electrode tube. As a rough approximation, one might assume that when their potentials are nearly equal, the currents to the plate and screen would be proportional to the ratio of the area of the openings in the screen to the area subtended by its lateral wires. Also, it might be expected that this ratio would increase slightly with increasing voltage of the plate with respect to the screen, because of a tendency of the plate to pull more electrons through the screen. This effect should be less for very fine mesh screens than for coarse ones.

From this simple theory, one would expect the plate current-plate voltage characteristics to be very flat for plate potentials higher than

the screen potential. Consequently, the plate conductance, given by

$$S_p = \frac{\partial I_p}{\partial E_p} = \frac{1}{R_p}$$

is a very small quantity; and the plate resistance, given by

$$R_p = \frac{1}{\frac{\partial I_p}{\partial E_p}}$$

is a very large quantity compared with its value in triodes, and increases with the fineness of mesh of the screen grid. The plate-current curves of Fig. 2 are seen to be in general accord with this simple theory at the higher values of plate voltage, although they are not quite as flat as might be expected from the theory. This, and the rapid falling off in the vicinity of 100 volts, will be discussed later.

Taken alone, the extremely high resistance of a screen-grid tube would seem to be a very serious disadvantage. From equation 14, the tube may be considered as analogous to a generator the electromotive force of which is $\mu e_g$ and the internal resistance of which is $R_p$, working into an external load resistance $R$—a generator with extremely high internal resistance. Why this is not fatal to the usefulness of the tube will be pointed out later.

While the plate current in a screen-grid tube is nearly independent of plate voltage for values of the latter higher than the screen potential, this obviously cannot hold at low values of plate voltage. At zero plate voltage, the plate current must be zero. At this point the screen collects the entire space current and $I_s = I_t$. As the plate potential increases from zero, the plate current would be expected to rise rapidly, with a corresponding drop in the screen current, as the plate collects more and more of the electrons passing through the screen. However, two factors tend to prevent an abrupt rise in the plate current to its nearly constant value when the plate becomes slightly positive. The first of these is space charge in the region closely adjacent to the plate produced by the electrons that pass through the screen, reach zero velocity in the region adjacent to the plate, and return to the screen. Some of them may perform several oscillations to and fro through the screen before being captured by it. This space-charge effect is largely masked by the more important second factor which is the deflection of the majority of the electrons from their normal paths by the intense electric fields about the lateral wires of the screen. This results in large differences in the components of velocity

of the electrons normal to the surface of the plate and, consequently, in the distance to which they approach the plate in their trajectories before being turned back to the screen. As a result, the plate must become positive by several volts with respect to the cathode before it captures substantially all of the electrons that pass through the screen.

From this simple theory, the plate-current and screen-current curves would be expected to have the form shown by the ideal curves of Fig. 2. Obviously, the screen-current curves must be complementary to the plate-current curves since the sum of the two currents is substantially constant.

The difference between these ideal curves and the actual characteristics, in the region extending from a few volts to potentials somewhat higher than the screen voltage, is attributed to the phenomenon of secondary electron emission. When electrons strike a metal surface with velocities equivalent to more than a few volts, other electrons, known as secondary electrons, are liberated from the surface. The number of electrons so liberated varies not only with the velocity of the primary bombarding electrons, but also with the character of the metal surface, the amount of adsorbed gases and other materials on the surface, and other factors. The number of such electrons leaving the surface may even exceed the number of primary electrons striking it, in which case the net current to the metal surface is negative. The velocity of the secondary electrons varies greatly. A very few have velocities approaching that of the primary electrons. The great majority, however, have low velocities equivalent to only a few volts.

In the screen-grid tube, an appreciable number of secondary electrons is liberated from the plate at potentials between 5 and 10 volts, and they increase in number with plate voltage. For plate potentials lower than the screen potential, in this case 75 volts, the secondary electrons from the plate are drawn to the screen, thus increasing the screen current by the amount the plate current is decreased. When the plate reaches a potential equal to that of the screen, secondary electrons no longer can escape from the plate to the screen, except those emitted with appreciable velocities; consequently, the plate current rises rapidly to its normal value.

At plate potentials higher than the potential of the screen, secondary electrons emitted from the latter are drawn to the plate. Consequently, in this region the plate current is slightly higher than it would be in the absence of secondary electron emission from the screen. The gradual rather than abrupt rise in the plate current curves at 75 volts is attributed primarily to the distribution of velocities with which the secondary electrons are emitted; to a lesser extent, it is dependent also

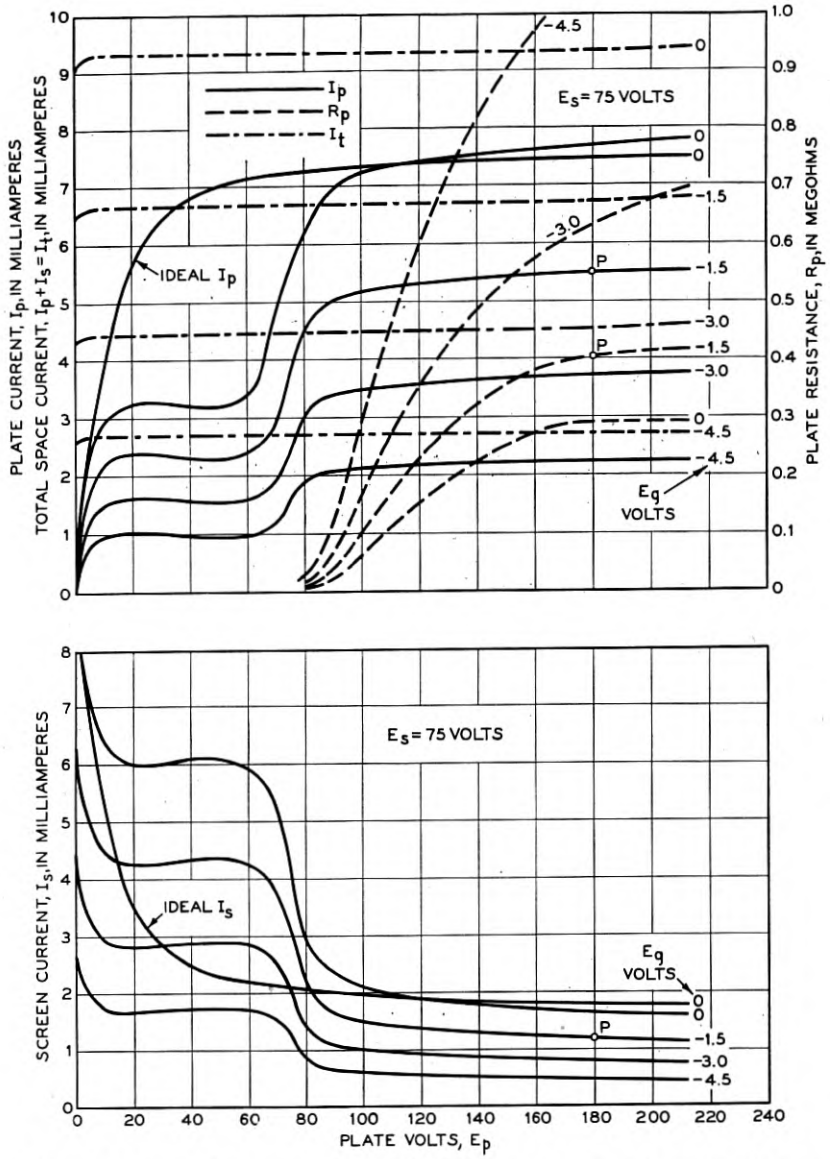


Fig. 2—Characteristics of a screen-grid tetrode. The curves show the plate current, $I_p$, screen current, $I_s$, total space current, $I_t = I_p + I_s$, and plate resistance, $R_p$, as functions of the plate voltage. Control-grid voltage, $E_g$, is indicated on the curves. Screen voltage, $E_s$, is maintained constant at 75 volts throughout.

on the combined effects of space charge, intensity and distribution of the field at the surface of the screen wires.

Normal operating conditions for the screen-grid tube the characteristics of which are shown in Fig. 2 are: $E_p = 180$ volts, $E_s = 75$ volts, and $E_g = -1.5$ volts. At this operating point $P$ in Fig. 2, the plate current is 5.5 ma and the plate resistance is 400,000 ohms. The operating range is confined to the flat portion of the characteristics. If the tube is operated with plate-voltage swings sufficiently large that the instantaneous values of the plate potential extend into the region of rapidly falling plate current, serious distortion of the output results. This is a serious limitation in screen-grid tetrodes, because it requires that the operating plate potential be much higher than the screen potential. How this limitation may be removed by the introduction of an additional electrode in the tube will be shown later.

Curves showing the variation of plate resistance with plate voltage are shown in Fig. 2. The plate resistance decreases with plate voltage, falling off very rapidly as the plate voltage approaches the screen voltage. The ordinates of the plate-resistance curves give a measure of the flatness of the plate-current curves. That the latter are not as flat as the total-space-current curves, thus resulting in values of plate resistance approaching infinity, is attributed to two factors: secondary electron emission from the screen, and an increasing ratio of plate current to screen current with increasing plate voltage. The increasing percentage of the primary space current drawn to the plate with increasing plate voltage is an involved and undetermined function of several factors including: the ratio of the openings in the screen grid to the total conducting area subtended by it, the intensity and distribution of the field at the screen grid, the velocity and directional distribution of the electrons arriving at the screen grid, and space-charge effects in its vicinity.

Thus, there is the interesting situation in screen-grid tubes that, for any given set of operating voltages, the magnitudes of the plate resistance and amplification factor are determined largely by factors not directly determined by the geometry and design of the tube. This is quite different from triodes in which the plate resistance and amplification factor both are determined directly by geometrical dimensions and the arrangement of the electrodes.

Obviously, the number of electrical parameters in vacuum tubes increases rapidly with the number of electrodes. The curves of Fig. 2, which correspond to the usual plate current-plate voltage characteristics for a triode, were obtained with the screen maintained at a constant potential of 75 volts as a fixed parameter. To obtain a complete

charting of the characteristics would require several such families of curves taken with different screen voltages. One such additional family of characteristics, in which $E_s$ is maintained at 90 volts, is shown in Fig. 3.

Since the plate in a screen-grid tube has practically no effect on the total space current, so far as consideration of the latter is concerned, one may regard the plate as being removed from the structure and consider only the remaining elements. The cathode, control grid, and screen then may be regarded as constituting an ordinary triode. By maintaining the screen at a positive potential, it is enabled to perform the function usually performed by the plate of a triode, *viz.*, that of supplying the positive field necessary to produce the flow of space current against the opposing resistance due to space charge. This function does not interfere with its screening action, so long as its potential is not allowed to vary. In fact, there is an actual gain in efficiency, as will be shown later, in having the electrode that provides the main driving field for the space current maintained at a fixed potential instead of varying over the operating cycle, as it must in the triode.

Usual design principles and equations applicable to triodes are also applicable to screen-grid tetrodes. By making the spacings between the electrodes small, particularly that between the cathode and control grid, a high transconductance can be obtained. Using the subscript $t$ to designate total space current, the transconductance is given by $S_t = \frac{\partial I_t}{\partial E_g}$ and has the same value as that for a triode of the same dimensions. Let $f$ represent the fraction of the total space current that passes through the openings in the screen and is collected by the plate, which is assumed to be at a higher potential than the screen. By proper design, $f$ can be made large, say from 0.7 to 0.9, in the normal operating range. The transconductance from the control grid to the plate is given by

$$S_m = \frac{\partial I_p}{\partial E_g} = f \cdot \frac{\partial I_t}{\partial E_c},$$

which is 0.7 to 0.9 of the normal value of transconductance for the three-electrode tube.

The amplification factor, $\mu$, for the tetrode, is given by

$$\mu = \frac{\frac{\partial I_p}{\partial E_g}}{\frac{\partial I_p}{\partial E_p}} = S_m \cdot R_p. \tag{17}$$

Since the transconductance, $S_m$, has a value not greatly different from the normal value for a triode, and since $R_p$ is very large, $\mu$, which is proportional to $R_p$, also must be very large. Referring to equation 14

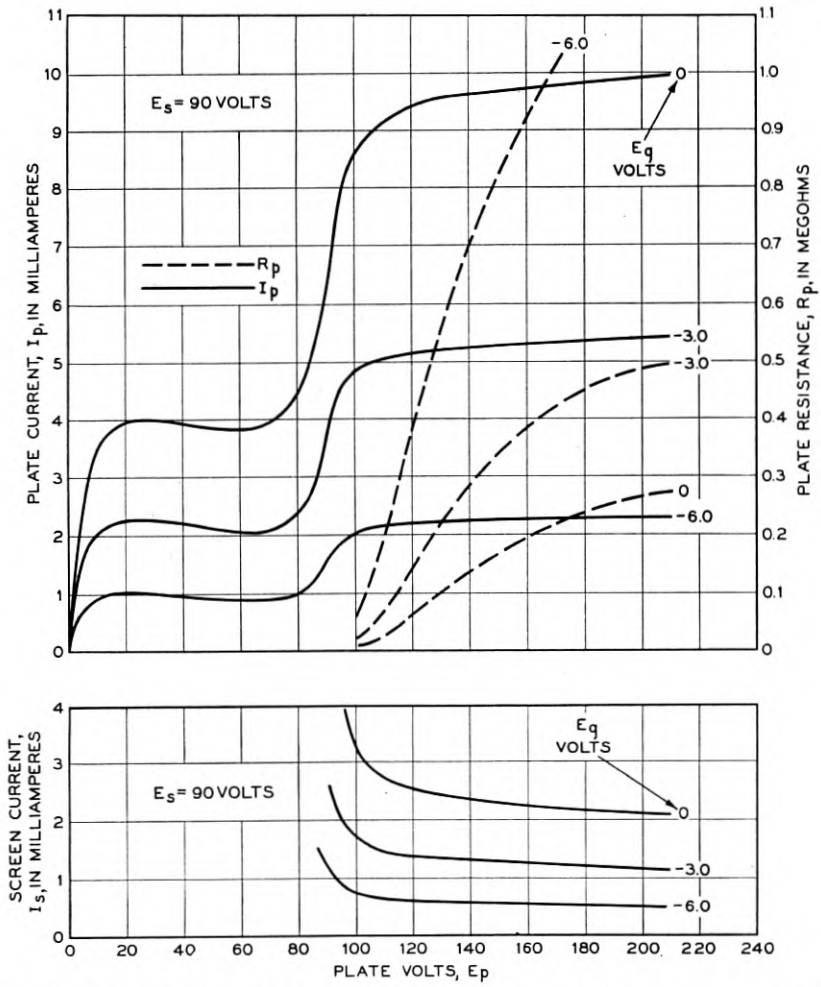


Fig. 3—Characteristics of a screen-grid tetrode similar to those shown in Fig. 2 except that $E_s$ = 90 volts.

and its analogy with the generator equation, it may be seen that, despite the fact that the screen-grid tube considered as a generator has a very high internal resistance, it gives normal output because it is provided also with a very large electromotive force, $\mu e_g$.

From this analysis of the characteristics of the screen-grid tube,

based upon the simple theory of triodes, one would expect that families of characteristics corresponding very closely with those of ordinary triodes would be obtained if the total space current, or the plate current,

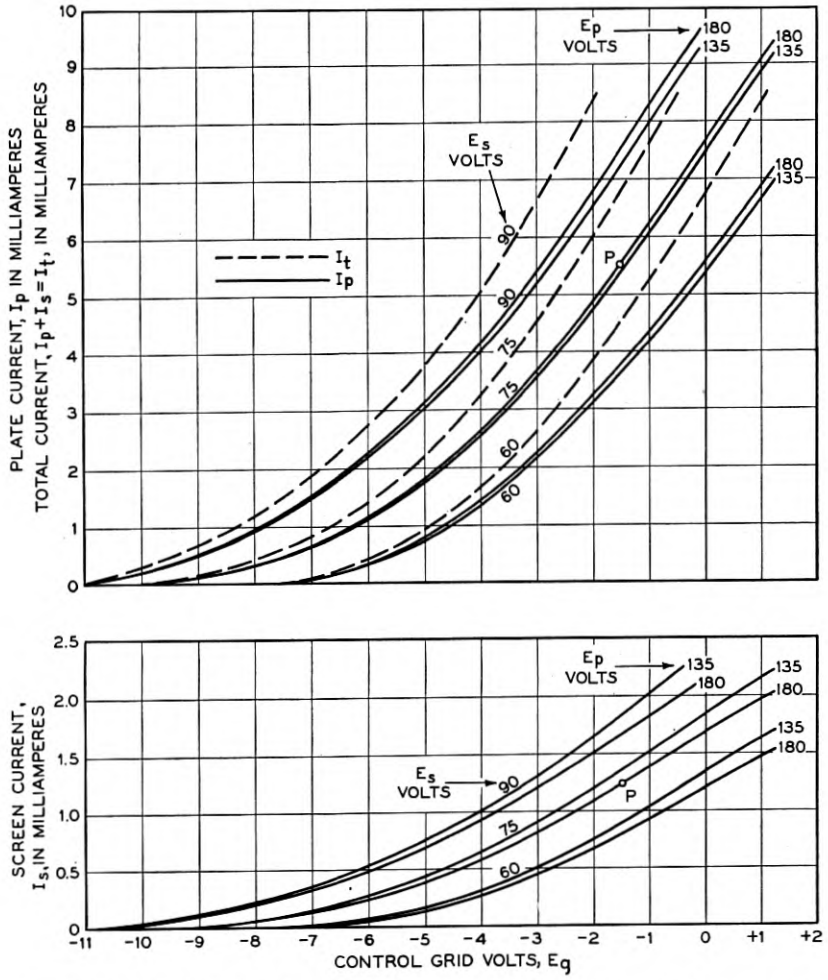


Fig. 4—Characteristics of a screen-grid tetrode. Plate current, $I_p$, screen current, $I_s$, and total space current, $I_t$, as functions of the control-grid voltage, with various values of plate voltage and screen voltage as parameters.

were plotted as functions of the control-grid voltage and screen-grid voltage, with the plate maintained at any fixed potential higher than the screen potential. It would be expected further that the family of curves showing total space current would remain practically invariant

with variations in plate voltage. The plate-current family of characteristics should show only small variations with plate voltage (so long as $E_p$ is higher than $E_s$), depending on the magnitude of the change in the ratio of $I_p$ to $I_s$.

Families of such characteristics are shown in Fig. 4. They are seen to correspond very closely indeed with similar characteristics for a triode. The $I_t$ curves vary so little with plate voltage that the families of characteristics taken at $E_p = 135$ volts and $E_p = 180$ volts, coincide within the breadth of the curves. The plate current curves show a small variation with $E_p$, as was discussed previously.

In Fig. 5 the transconductance and amplification factor are shown as functions of the plate voltage for four different values of grid bias and with the screen potential maintained constant at 75 volts. At the normal operating point $P$, the transconductance is 1375 micromhos and the amplification factor is 550. The shape of these curves is typical of that for screen-grid tubes and pentodes. Throughout the normal operating range, the transconductance curves have about the same degree of flatness as the plate-current curves. This is to be expected from consideration of the plate-current curves of Fig. 4. Since these curves change only slightly with variations in plate potential, their slopes or transconductance values also change but slightly. The amplification-factor curves are very similar in form to the plate-resistance curves of Fig. 2. This follows at once from equation 17, for since $S_m$ remains nearly constant with variations in $E_p$, $\mu$ must vary in the same manner as $R_p$.

In Fig. 6 the transconductance, plate resistance, and amplification factor are shown as functions of grid voltage for three different values of the screen voltage and with the plate voltage maintained constant at 180 volts. These curves are also typical of those found for several multi-electrode tubes. The transconductance curves agree in form with those for a triode, as would be expected from the plate-current curves of Fig. 4. The plate-resistance curves are similar in form to those for a triode, but rise more rapidly with increasing negative grid bias. The amplification-factor curves, however, are entirely different in form from those for a triode, since they rise with increasing negative grid bias, whereas in triodes the amplification factor decreases with increasing negative grid bias. The reason for this difference is that the plate resistance in many multi-electrode tubes, over the normal range of operation, increases more rapidly with decreasing plate current than in triodes. At sufficiently large negative values of grid bias and very low plate currents, the amplification-factor curves frequently reach maxima and then fall rapidly as they do in triodes.

In the operation of triodes as voltage amplifiers, it is usually possible to have the external load resistance large with respect to the plate resistance. Consideration of equation 15 shows that this results in a

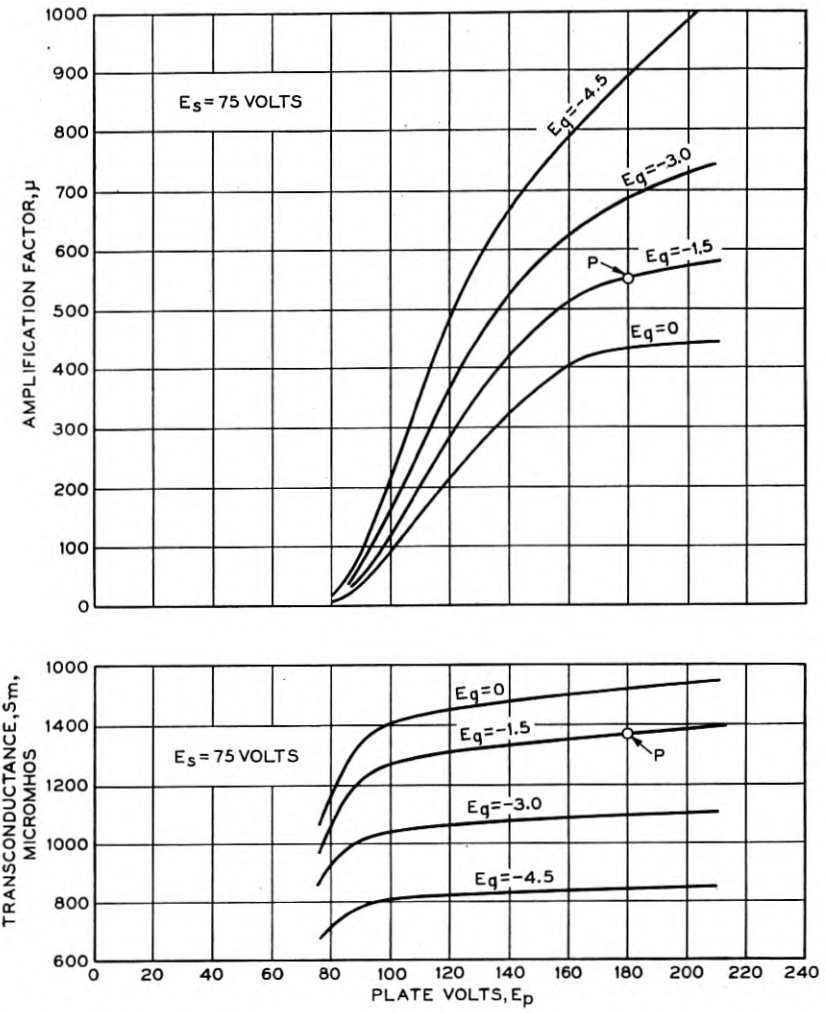


Fig. 5—Amplification factor and transconductance of a screen-grid tetrode as functions of the plate voltage.

voltage amplification ratio approaching the amplification factor. In the operation of screen-grid tubes and conventional pentodes as well, which have high plate resistances, it is usually necessary that the load resistances be much smaller than the plate resistances. This limitation

is imposed by circuit coupling requirements and sometimes by restrictions on the permissible harmonic content in the power output. This limitation results in an amplification ratio that is much smaller than

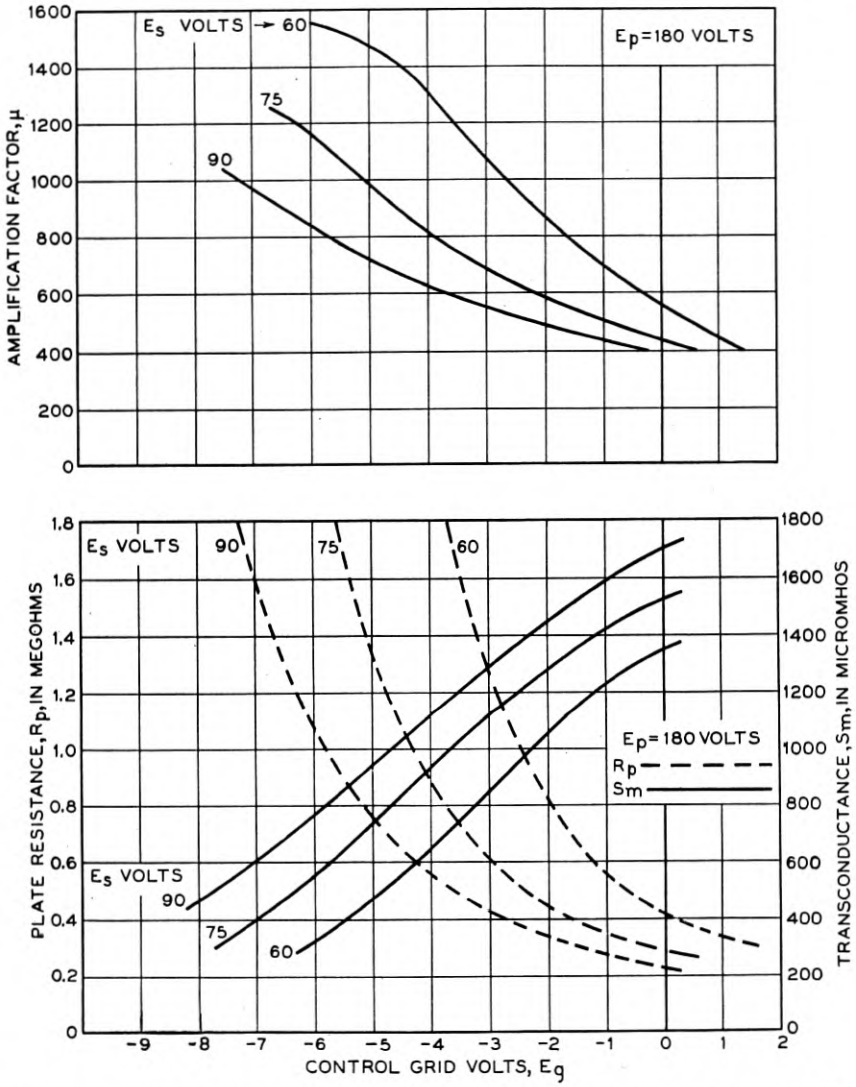


Fig. 6—Amplification factor, transconductance, and plate resistance of a screengrid-tetrode as functions of the control-grid voltage.

the amplification factor; but even so, a considerably higher amplification ratio usually can be obtained than with triodes. This is apparent

from consideration of equation 16 which may be written in the form,

$$A_v = S_m \cdot \frac{R}{1 + \frac{R}{R_p}} \cdot \qquad (18)$$

Since in screen-grid tubes and pentodes, $R/R_p$ is small, the voltage amplification is given approximately by

$$A_v = S_m \cdot R. \qquad (19)$$

Now, in a comparable triode the transconductance will have a value comparable with $S_m$, and it may be assumed that a load resistance could be used comparable with $R$. However, the ratio $R/R_p$ for the triode is not small and may be greater than unity. Hence, from equation 18, the amplification obtained with the triode is correspondingly smaller.

It has been shown that the presence of the electrostatic screen in the screen-grid tetrode results in a high plate resistance and a high amplification factor, but that the ratio of $\mu$ to $R_p$ (transconductance) remains normal. Such tubes yield high amplification and, with suitable associated circuits, are relatively free from feed-back; but they are limited in their range of operation because of the fold in the plate current-plate voltage characteristics resulting from secondary emission from the plate.

## Power Pentodes

In order to deliver a large power output, a vacuum tube must be capable of large variations in plate current and plate voltage from their normal operating values. Both of these conditions are fulfilled by the power pentode. The arrangement of the electrodes, shown schematically in Fig. 7, corresponds to that in a screen-grid tube except that an additional grid, $g_3$, is inserted between the plate and screen grid, $g_2$. As in the screen-grid tube, the first grid, $g_1$, has a negative voltage applied to it and acts as the control element. The second grid, $g_2$, is maintained at a fixed positive potential, $E_s$, and provides the main driving field for the space current.

As in the screen-grid tube, the total space current is determined almost wholly by the geometrical dimensions and spacings of the cathode, $g_1$ and $g_2$, and by the voltages applied to these electrodes. Consequently, in the design of this portion of the structure, the same considerations apply as in the design of an ordinary triode to deliver large power output. By making the inner grid comparatively coarse and by designing the second grid to operate at comparatively high potentials, a structure having a low amplification factor is obtained

which draws a large space current from the cathode at a control-grid bias sufficiently negative to permit relatively large swings of the control-grid voltage. In pentodes designed to operate at low frequencies, screening between the plate and control grid is unimportant;

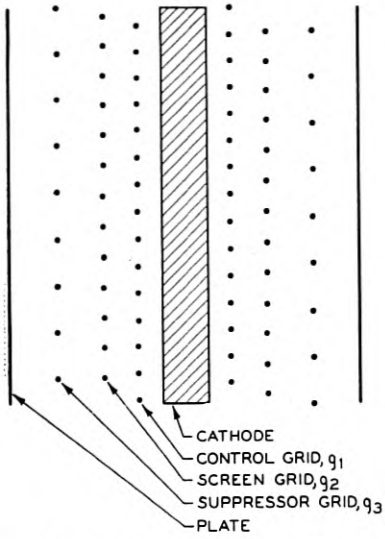


Fig. 7—Schematic diagram showing the arrangement of electrodes in a power pentode.

hence, the second grid can be comparatively coarse, thus permitting as large a portion as possible of the space current to pass through it to the plate.

To permit the largest possible swings in plate voltage, it is necessary to remove the "fold" in the plate-current characteristics, caused by secondary electrons emitted from the plate. This is accomplished by the insertion of a third grid, $g_3$, between the plate and second grid, $g_2$. This grid, known as a suppressor grid, must be maintained at a lower potential than the lowest instantaneous potential reached by the plate, and is usually maintained at the cathode potential by connecting it to the cathode inside the tube. The suppressor grid exerts a retarding force on the primary electrons flowing toward it from the cathode, but, because of its coarse structure, all but a small fraction succeed in passing through it and are accelerated again, finally reaching the plate with the same velocity they would have if the suppressor grid were absent. On the other hand, secondary electrons emitted either by the plate or screen grid find themselves in a retarding field, which they are unable to traverse because of their low velocity, and are constrained to return to the electrode from which they came.

Plate current-plate voltage characteristics and screen current-plate voltage characteristics for a power pentode of the indirectly heated cathode type, are shown in Fig. 8. The secondary emission "fold"

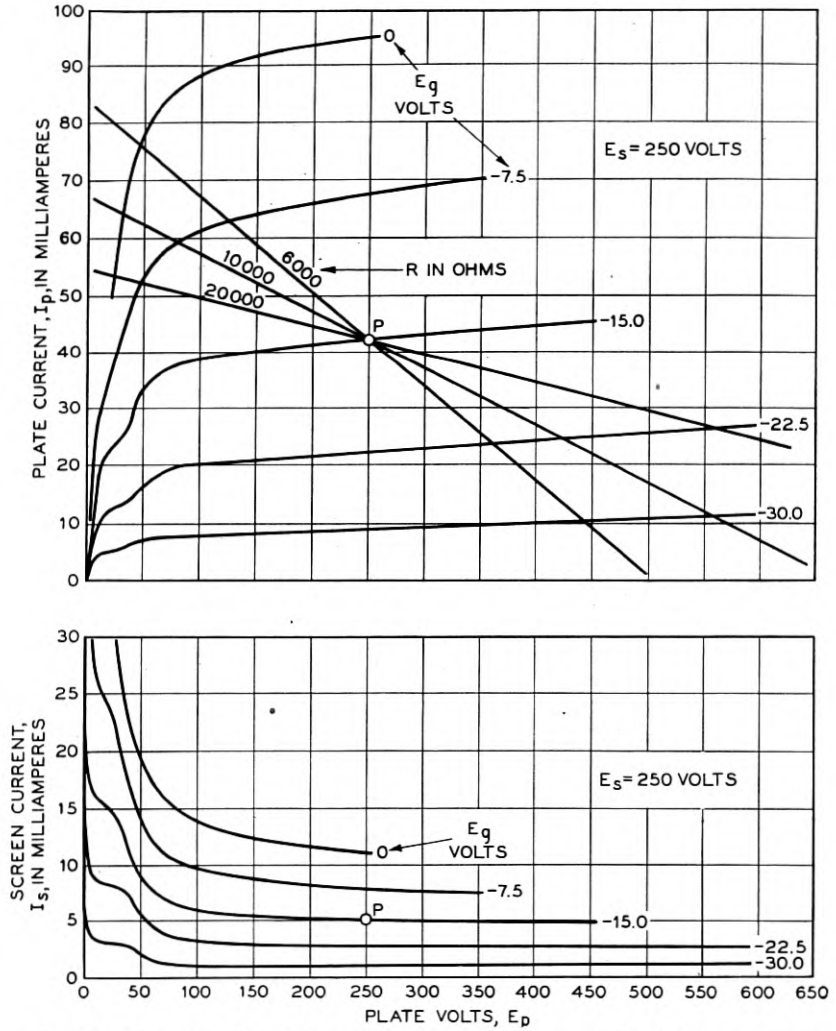


Fig. 8—Characteristics of a power pentode. Plate current, $I_p$, and screen current, $I_s$, as functions of the plate voltage with various values of the control-grid voltage, $E_g$, as parameters. Screen voltage, $E_s$ = 250 volts throughout. Load lines are shown for resistance loads of 6000, 10,000, and 20,000 ohms.

in the characteristics is almost completely eliminated by the suppressor grid.

The effectiveness of the suppressor grid is shown by the charac-

teristics of Fig. 9. These curves were obtained from a tube of the same type as that for which the characteristics are shown in Fig. 8. One set of curves was obtained with the suppressor grid operating in the normal manner. The other curves were obtained with the suppressor grid tied to the screen grid and maintained at a positive

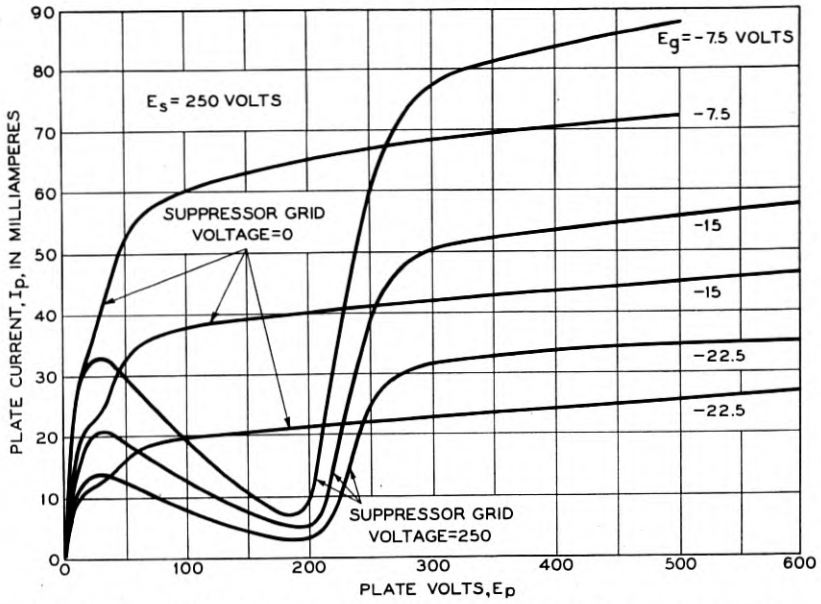


Fig. 9—Curves showing the effect of the suppressor grid in a power pentode.

potential of 250 volts. In this latter case, the number of secondary electrons escaping from the plate is practically the same as if the suppressor grid were removed from the tube. The presence of the suppressor grid not only permits the plate to swing to very much lower potentials than otherwise would be possible, but also permits the plate and screen to operate at the same potential, which is 250 volts in this case.

The characteristics of Fig. 8 closely approach the form that would be expected from simple theory. As is usual with power pentodes, the plate-current curves are not quite as flat as those for screen-grid tubes. This is because of the more open character of the grids which permits the plate to have a slightly greater effect on the magnitude of the space current. This is evidenced also by the tendency of the curves to turn up at the higher plate voltages.

The normal operating point for this tube is at point $P$ in Fig. 8, at which the plate and screen potentials are both 250 volts and the control

grid potential is $-15$ volts. Under these conditions, the average characteristics are: $I_p = 42$ ma, $I_s = 5$ ma, $\mu = 156$, $R_p = 52{,}000$ ohms, and $S_m = 3000$ micromhos.

Curves showing the amplification factor, $\mu$, the plate resistance, $R_p$, and the transconductance, $S_m$, for several different values of grid bias, are shown in Fig. 10. They correspond in general form to those

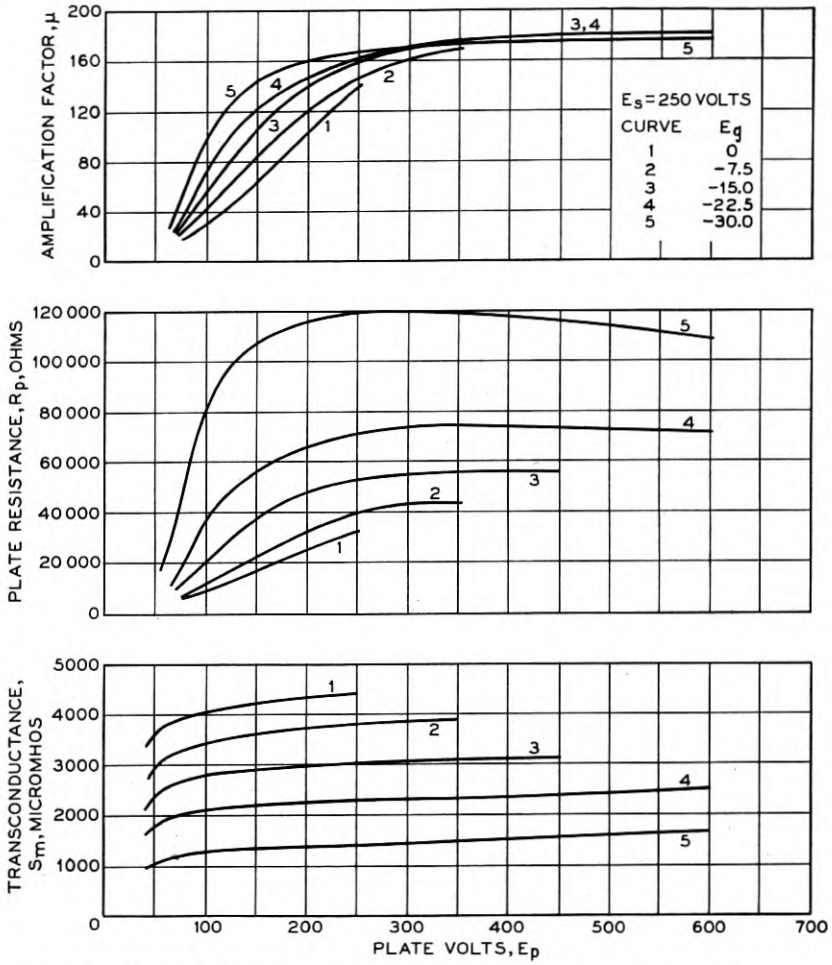


Fig. 10—Amplification factor, plate resistance, and transconductance of a power pentode as functions of the plate voltage, with various values of control-grid voltage as parameters. Screen voltage, $E_s$, is maintained constant at 250 volts.

shown in Figs. 3 and 5 for the screen-grid tube. The maxima in some of the plate-resistance curves result from the fact that the corre-

sponding plate-current curves turn up at the higher plate potentials and thus have points of inflection.

In Fig. 11, the plate current is shown as a function of control-grid

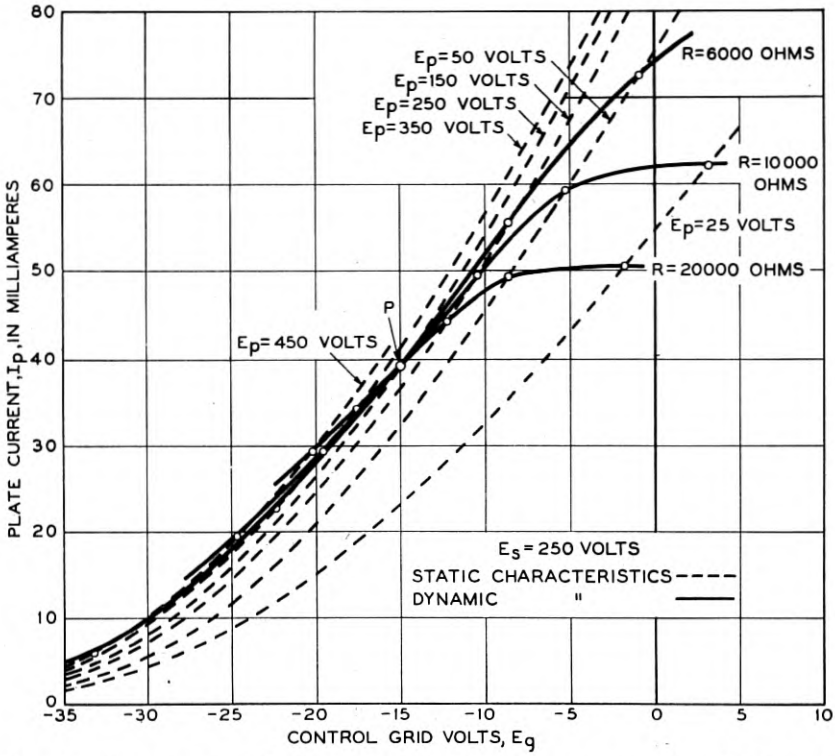


Fig. 11—Plate current-grid voltage characteristics of a power pentode, with dynamic characteristics for resistance loads of 6000, 10,000, and 20,000 ohms.

voltage for various values of the plate voltage, with the screen potential constant at 250 volts. The curves are very similar to corresponding ones for the screen-grid tube shown in Fig. 4. Here, curves are included at such low plate voltages that a large falling off in plate current occurs.

In Fig. 12, the output power in watts, and the second and third harmonics, expressed in decibels below the fundamental, are shown as functions of peak volts for a sinusoidal input applied to the grid. These data were obtained under the normal operating conditions previously given, and with the indicated load resistances. Since the curves are typical in form of those obtained in several types of multi-electrode tubes, it will be of interest to examine them in some detail.

Except at the higher inputs, the power output increases continuously with load resistance over the load range considered. This is to be expected, since the highest load resistance of 12,000 ohms is much

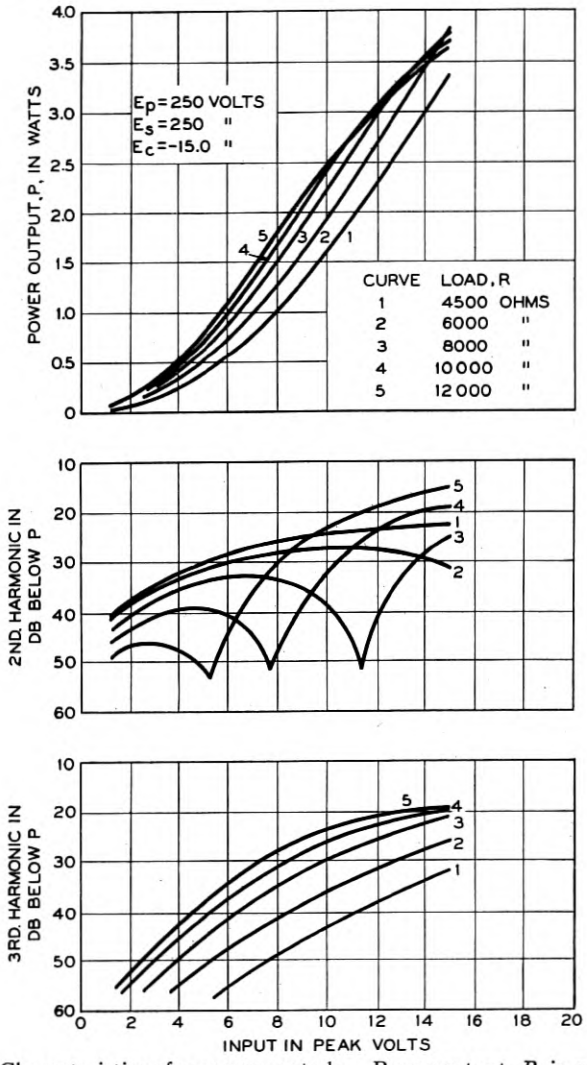


Fig. 12—Characteristics of a power pentode. Power output, $P$, in watts, second and third harmonics in decibels below the fundamental, as functions of the peak input when a sinusoidal voltage is applied to the control grid.

smaller than the plate resistance of the tube, which is 52,000 ohms. For small inputs, the maximum power output would be obtained, as it is in triodes, when the load resistance is equal to the plate resistance of

the tube. The decrease in power output with increasing load resistance, at the higher inputs, results from the progressive turning over of the dynamic characteristics as shown by the curves of Fig. 11.

The second harmonic decreases continuously with increasing load resistance at small inputs. At large inputs, it decreases at first, then increases with increasing resistance. At very low resistances, the second harmonic increases continuously with input. At higher load resistances, it rises to a broad maximum, then falls to a very sharp minimum, after which it again rises rapidly with increasing input. The explanation of these phenomena will be given later.

The third harmonic increases continuously both with increasing input and with increasing load resistance. At load resistances that are small compared with the plate resistance of the tube, it rises to a higher level than the second, over a certain range of input. Relatively high levels of third harmonic are characteristic of pentodes, screen-grid tubes, and also of some other types of multi-electrode tubes. For example, from the curves of Fig. 12, with a load resistance of 6000 ohms and an input of 15 peak volts, the power output is 3.8 watts with the volume level of the second and third harmonics 31 db and 26 db, respectively, below that of the fundamental. At an input of 10 peak volts, the volume level of the second harmonic rises to 27 db and of the third harmonic falls to 36 db below that of the fundamental.

These results are quite different from those obtained with triodes where the third harmonic is, in general, 10 db or more below the second. Furthermore, in triodes, both the second and third harmonics decrease continuously, as a rule, with increasing load resistance. There are exceptions to this, however, where the third harmonic curves show minimum points or cusps, similar to those shown by the second harmonic curves for pentodes.

The reason for the relatively large harmonic content in the output of pentodes is apparent from consideration of the load lines drawn through the operating point $P$ in Fig. 8. At the lower values of plate voltage, the load lines cut across the rapidly descending portions of the plate-current characteristics. This effect is more marked and begins at more negative values of grid bias as the load resistance increases and the slope of the load lines becomes correspondingly less. The effect of this is to produce current variations through the external load resistance that are not proportional to the variations in grid voltage, thus resulting in distortion of the output.

The character of the distortion is made clearer by reference to the dynamic characteristics of Fig. 11, for resistance loads of 6000, 10,000, and 20,000 ohms. All the curves show a flattening out at the top which

increases progressively with load resistance. It readily is shown that such dynamic characteristics, having points of inflection at which the curvature changes sign, give rise to peculiarities in the harmonic output.

If such a characteristic be expressed by a power series in terms of grid-voltage variations from the operating point $P$, there usually is found a relatively large contribution by third and higher odd-power terms the coefficients of which are predominantly negative in sign. Since odd-power terms yield odd harmonics, this accounts for the relatively high levels of third harmonic at input voltages sufficiently large that the flat portion of the characteristic is traversed. However, positive and negative signs are about evenly divided among the coefficients of the even-power terms, which yield even harmonics. At some value of the input voltage, which varies with the load resistance, the contributions to the second harmonic by positive and negative terms are approximately equal, resulting in a very small value of this harmonic. This accounts for the cusps in the second harmonic curves of Fig. 12. For inputs less than the value at the cusp, the contribution of positive terms (largely the second-power term) prevails over that of negative terms, while at higher inputs the reverse is true. Consequently, there is a reversal in the phase of the second harmonic at the cusp.

If the point of inflection were at the operating point $P$ in Fig. 11 and, if the dynamic characteristic were symmetrical about $P$, then only odd-power terms would appear in the equation of the curve. Consequently, in this special case, even harmonics would vanish from the output and only odd harmonics would remain.

Plate-circuit efficiency of pentodes is higher than that usually found in triodes. The underlying physical reasons for this difference are as follows: In the triode, the plate simultaneously performs two functions. First, it is the element in the output circuit whose fluctuating potential is impressed across the load resistance. Second, assuming that the grid potential is not positive at any time, the plate is the only positive electrode providing the necessary driving force for the space current. These two functions militate against each other to a certain extent, for, as is evident from consideration of the dynamic characteristic of either a triode or a pentode, the plate voltage reaches its minimum value at the instant when the plate current reaches its maximum value. This minimum voltage, which must be sufficiently large to draw the peak current through the tube, is a very substantial fraction of the operating plate voltage, particularly when the latter is comparatively low.

In pentodes and in some other multi-electrode tubes, the positive grid, maintained at a constant voltage, provides the necessary driving

force for the space current, thus relieving the plate from performing this function. Consequently, the plate is free to swing to lower voltages than otherwise would be possible, which results in a corresponding increase in efficiency.

In Table I, similar data are shown for typical pentodes and triodes.

TABLE I

COMPARATIVE DATA FOR TYPICAL TRIODES AND PENTODES

| | Triodes | | | Pentodes | | |
|---|---|---|---|---|---|---|
| | *A* | *B* | *C* | *D* | *E* | *F* |
| Plate voltage, $E_p$ | 200 | 250 | 325 | 180 | 250 | 250 |
| Screen voltage, $E_s$ | | | | 180 | 250 | 250 |
| Grid voltage, $E_g$ | −45 | −50 | −68 | −18 | −16.5 | −15 |
| Plate current, $I_p$, ma | 45 | 34 | 60 | 14.5 | 34 | 42 |
| Screen current, $I_s$, ma | | | | 2.8 | 6.5 | 5 |
| Transconductance, $S_m$, micromhos | 2,810 | 2,175 | 5,200 | 1,050 | 2,200 | 3,000 |
| Amplification factor, $\mu$ | 2.9 | 3.5 | 3.8 | 105 | 220 | 156 |
| Plate resistance, $R_p$, ohms | 1,030 | 1,610 | 730 | 100,000 | 100,000 | 52,000 |
| Load resistance, $R$, ohms | 2,060 | 3,900 | 2,750 | 12,000 | 7,000 | 6,000 |
| Input, peak volts | 45 | 50 | 68 | 18 | 16.5 | 15 |
| Power output, watts | 1.9 | 1.9 | 7.8 | 1.2 | 3.0 | 3.8 |
| Second harmonic, % | 7.4 | 5.7 | 5.0 | 5.0 | 2.3 | 2.8 |
| Third harmonic, % | 0.7 | 1.2 | 1.1 | 7.5 | 6.6 | 5.0 |
| Total harmonic, effective % | 7.45 | 5.9 | 5.1 | 9.0 | 7.0 | 5.7 |
| Plate efficiency, % | 21 | 22 | 40 | 46 | 35 | 36 |
| Plate-grid capacitance, $\mu\mu$f | 12 | 7.2 | 20 | 1.2 | 0.8 | 0.2 |
| Plate-ground capacitance, $\mu\mu$f | 3.2 | 3.0 | 5.5 | 7.8 | 8.5 | 15 |
| Grid-ground capacitance, $\mu\mu$f | 6.8 | 4.5 | 9 | 6.4 | 8.4 | 9 |

While the triodes and pentodes are not directly comparable with each other, the data are indicative of the differences between the two types of tubes. The chief differences between pentodes and triodes may be summarized as follows: (1) Pentodes yield higher gain and require correspondingly lower input voltages to drive them. (2) Pentodes have much higher amplification factors and plate resistances than triodes. The latter constitutes a handicap in coupling the tube to its circuit, particularly if transformer coupling is employed. (3) Pentodes yield high power output, generally at higher plate efficiency than triodes. (4) The harmonic content in the output of pentodes is high, the third harmonic being particularly high compared with its level in triodes. This requires that pentodes work into load impedances that are very low compared with the plate resistance. Ratios of $R$ to $R_p$ of 1/3 to 1/10 are common. (5) The plate-grid capacitance is much lower and the plate-ground capacitance is somewhat higher in pentodes than in comparable triodes.

## Space-Charge-Grid Pentodes

One limitation in the three-electrode tube and in the multi-electrode tubes considered thus far in this paper, is the resistance offered by space charge to the flow of space current. Tubes having so-called space-charge grids overcome this limitation to some extent by having a positive grid close to the cathode, which partially neutralizes the negative field very near the cathode surface due to space charge. A comparatively large current is drawn from the cathode by the space-charge grid. A portion of this current (usually about half of it) is collected by this grid, while the remaining portion passes through it and is acted on by the remaining elements of the tube.

The arrangement of the electrodes in a space-charge-grid pentode is shown in Fig. 13. The space-charge grid, $g_1$, is maintained at a

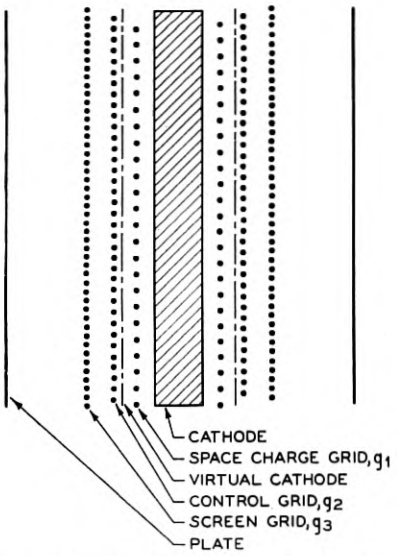


Fig. 13—Schematic diagram showing the arrangement of electrodes in a space-charge-grid pentode.

relatively low positive potential with respect to the cathode, usually in the range from 10 to 20 volts. The second grid, $g_2$, is the control grid and is maintained at a negative potential with respect to the cathode.

Ideally, at some cylindrical surface between these two grids (assuming the structure to be cylindrical) the electrons are retarded to nearly zero velocity forming a second space-charge region which may be regarded as a virtual cathode. Since this space-charge sheath is larger in area than the original cathode and is very close to the control grid, it

results in very large values of transconductance. Practically, the ideal condition is not fully realized, largely because velocity components other than radial are imparted to the electrons in passing through the space-charge grid. Consequently, these electrons reach any given cylindrical surface outside the space-charge grid with rather widely varying radial components of velocity. This, as will be seen later, places a rather serious limitation on the performance of such tubes.

The arrangement and functioning of the other electrodes in Fig. 13, outward from the virtual cathode, correspond with that of the screen-grid tetrode. The screen grid, $g_3$, is maintained at a fixed positive potential necessary to accelerate the electrons from the region of the virtual cathode. The plate must be maintained at a potential higher than that of the screen for the same reason as in the screen-grid tetrode. It will be shown that the characteristics of this pentode correspond roughly with those of the screen-grid tube previously discussed.

If $g_3$ were omitted, the structure outside the virtual cathode would correspond to that of a triode and the characteristics in the resulting tetrode would correspond roughly to those of a triode. In Fig. 14, characteristics are shown for a pentode of this type, the cathode and general dimensions of which are the same as those of the power pentode, the characteristics of which were shown previously. The plate-current and screen-current curves are seen to correspond very closely with those previously shown for a screen-grid tube. The characteristics exhibit the same "folds" due to secondary electrons, although this portion of the characteristics is not shown. The net or space-charge-grid current, $I_n$, increases as the plate current decreases with increasing negative control-grid voltage. This is to be expected since, as the control grid becomes more negative and reduces the current passing through it to the plate, the excess current returns to the net rather than to the cathode as in the screen-grid tube.

If values of the amplification factor, plate resistance, and transconductance are plotted as functions of the plate voltage, families of curves are obtained similar in all respects to those for a screen-grid tube as shown in Figs. 3 and 5. These characteristics are not shown for this tube.

In Fig. 15, plate-current, screen-current, and net-current characteristics are shown as functions of the control-grid voltage with different values of the screen voltage as parameters. As in the case of the screen-grid tube, only a slight displacement of the characteristics results from variation of the plate voltage. It is of interest to note the high values of net current, particularly as the plate current drops toward zero with increasing negative grid bias. For example, if the operating point is

chosen at $P$ in Fig. 15, with a screen potential of 90 volts and a grid potential of $-2$ volts, the plate current, screen current, and net current are 24, 0.5, and 23 ma, respectively, making a total space

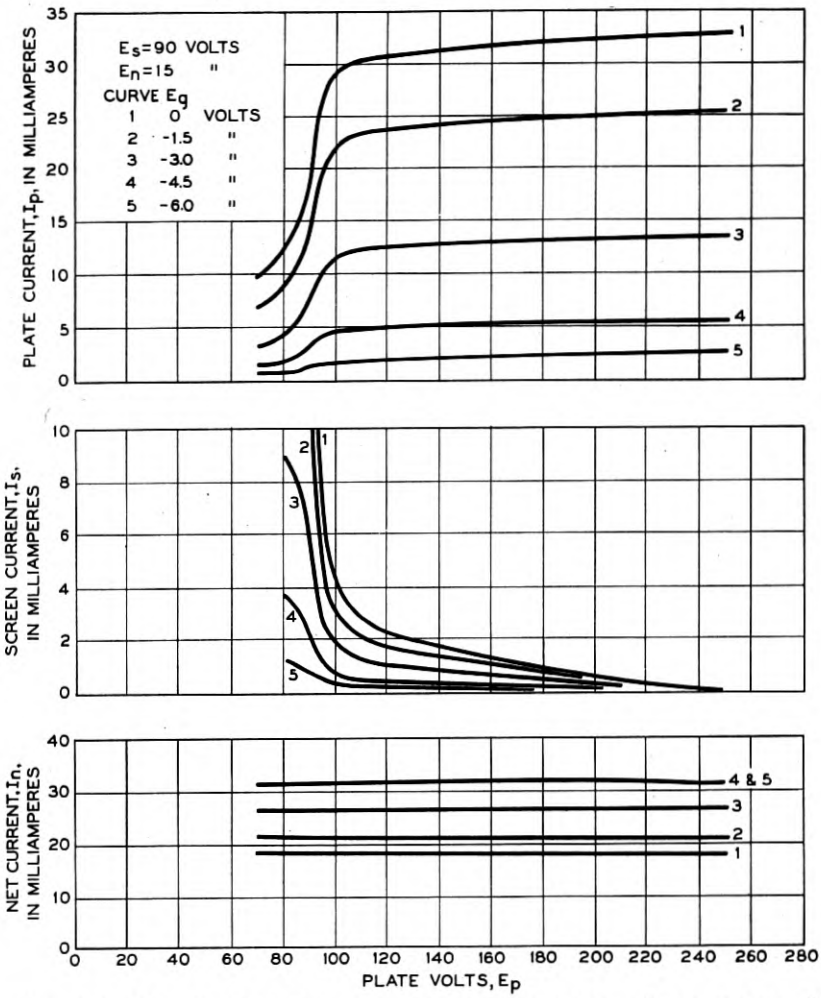


Fig. 14—Characteristics of a space-charge-grid pentode. Plate current, $I_p$, screen current, $I_s$, and net (space-charge-grid) current, $I_n$, as functions of the plate voltage with the indicated values of control-grid voltage as parameters. Screen-grid voltage, $E_s = 90$ volts. Net (space-charge-grid) voltage, $E_n = 15$ volts.

current of 47.5 ma drawn from the cathode, nearly half of which is collected by the space-charge grid.

Another point of interest in Fig. 15 is the flattening out of the plate-current characteristics at the higher values of plate current, in a manner

very similar to that exhibited by triodes of low cathode emission. In fact, this phenomenon is caused by the partial exhaustion of the space charge in the region of the virtual cathode. This is a fundamental characteristic of space-charge-grid tubes, caused largely by the

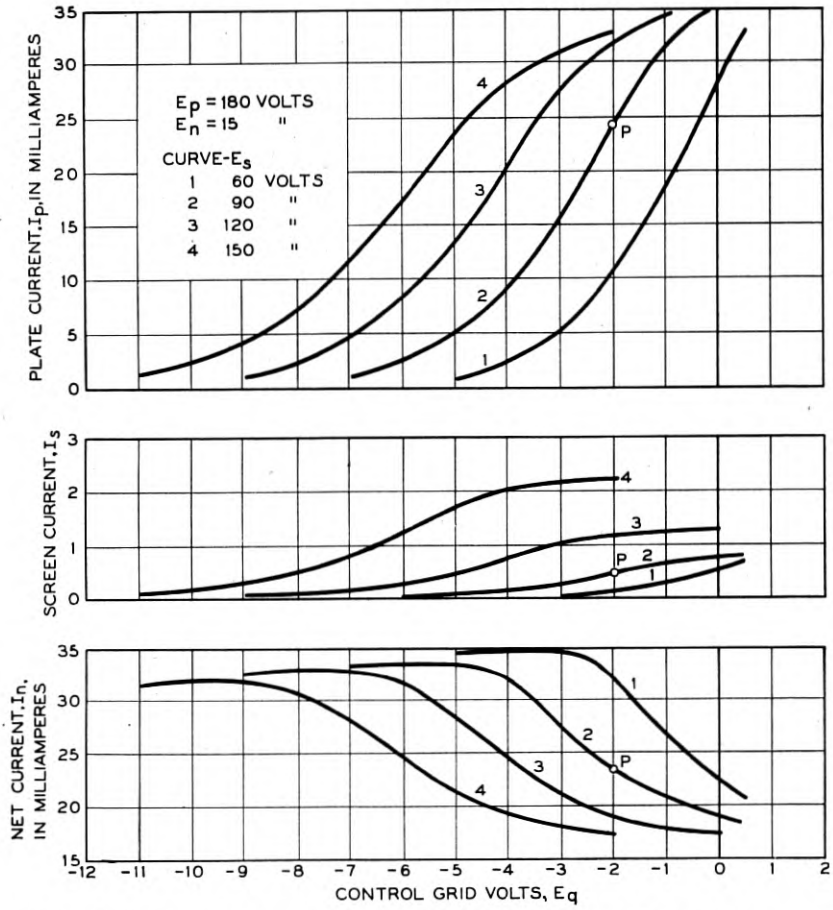


Fig. 15—Characteristics of a space-charge-grid pentode. Plate current, screen current, and net current as functions of the control-grid voltage, $E_g$, with the indicated values of screen voltage as parameters. Plate voltage, $E_p = 180$ volts. Net voltage, $E_n = 15$ volts.

imperfect character of the virtual cathode, which results from electrons entering the space-charge region with widely varying normal components of velocity. It constitutes a serious limitation on the practicable range of operation of such tubes, since it prevents the plate potential from swinging over a sufficiently large range to obtain a large power output without resulting in prohibitive distortion.

In Fig. 16, the amplification factor, plate resistance, and transconductance are shown as functions of control-grid voltage, with the same parameters as were used in obtaining the curves of Fig. 15. Of particular interest in these curves, are the unusually high values of

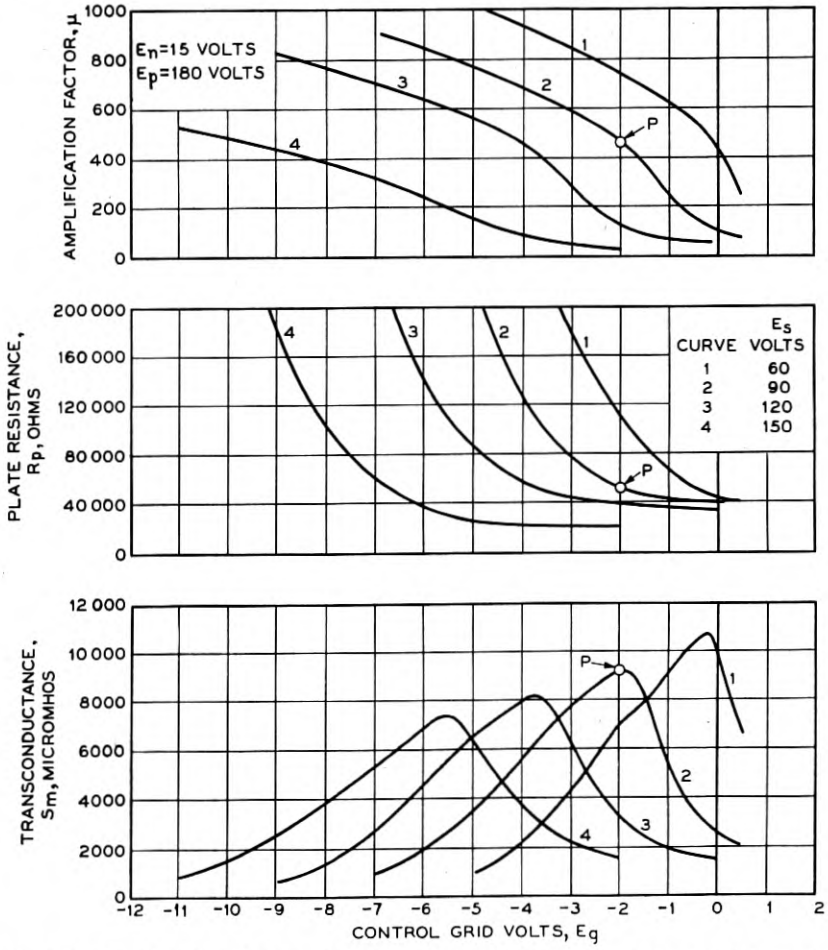


Fig. 16—Amplification factor, plate resistance, and transconductance of a space-charge-grid pentode as functions of the control-grid voltage.

transconductance for a tube of this size. For example, at the selected operating point $P$ it is 9200 micromhos. Also of interest is the peaked character of the transconductance curves. Since transconductance is defined by the slope of the plate current-grid voltage curves, which for this tube have points of inflection in them, maximum points must occur

in the transconductance curves. The rapid falling off in transconductance on either side of the maxima, with change in grid voltage, also indicates that a large amount of distortion would result if the tubes were operated with large swings in grid potential.

Space-charge-grid pentodes are capable, then, of yielding very high amplification because of their extraordinarily high transconductance; but, practically, they are limited to use as voltage amplifiers for fairly small inputs, since operation over the wide range necessary for large power output results in prohibitive distortion. These statements apply also to space-charge-grid tetrodes. These tubes have the advantage over ordinary triodes, however, of yielding high values of transconductance at comparatively low voltages applied to the plate.

## Co-Planar-Grid Tubes *

Thus far only multi-electrode tubes have been considered in which the electrodes are arranged in concentric order one about another. In the case of the power pentode it has been shown that this arrangement results in characteristics yielding comparatively high output power at high amplification and efficiency but with relatively large percentages of harmonics. It has also been pointed out that the high plate-circuit efficiency is largely due to the presence of a positive grid held at a fixed potential, which permits larger plate voltage swings than would otherwise be possible.

It is evident also from the underlying physical principles that in such a structure, with three concentric grids between the plate and cathode, even though the lateral spacings are fairly wide, enough shielding is interposed between the plate and cathode to make the plate resistance high. The question then arises: Is there any possible way of obtaining the advantages of a positive grid in tubes designed to operate at comparatively low plate voltages without the accompanying high plate resistance of the pentode structure?

This objective is accomplished reasonably well by a four-electrode structure in which the positive grid is placed in the same plane as the control grid. The arrangement of the electrodes in such a structure, which will be referred to as a co-planar-grid tube, is shown schematically in Fig. 17. The lateral wires of the two grids are asymmetrically arranged, the positive and negative lateral wires being arranged in pairs close together. The object of this arrangement may be described roughly as follows: By such an arrangement, when the control grid is very negative, the field about each negative wire very largely neutral-

* This section, "Co-Planar Grid Tubes," is an addition to the material published in *Electrical Engineering*.

izes the field due to its positive companion so that the latter becomes ineffective in drawing electrons away from the cathode. As the negative grid approaches zero potential it uncovers the positive grid electrically so that the latter becomes highly effective in drawing a large space current from the cathode at the moment when the peak current is drawn to the plate. This would not be so effectively accomplished if the lateral wires were symmetrically arranged.

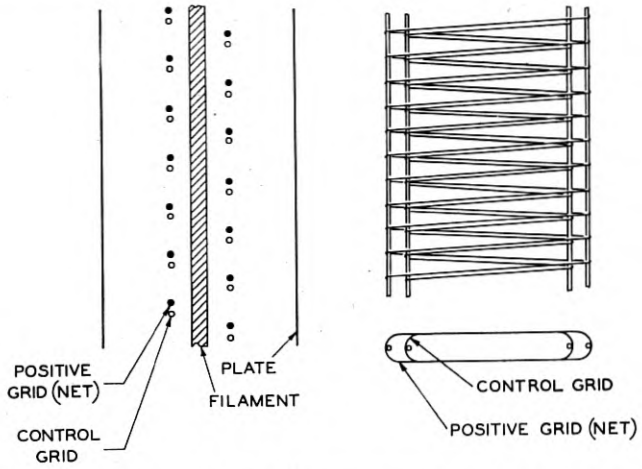


Fig. 17—Schematic diagram showing the arrangement of electrodes in a co-planar-grid tube.

In Fig. 18, characteristic curves are shown for a co-planar-grid tube designed to operate at a plate potential of 130 volts and at a plate current of 35 milliamperes. Each of the three groups of curves shows the plate current as a function of the control-grid voltage for plate voltages of 100, 130 and 160 volts, and with the positive-grid voltage maintained constant at the indicated value. It is evident from these curves that the primary effect of increasing the potential of the positive grid, or net, is to translate the group of plate-current curves to the left. By so doing it is obvious that larger control-grid swings are possible, at any given operating plate current, without the potential of the control-grid becoming positive at any time. This results in a comparatively large power output, to obtain which large grid swings are essential.

The lower curve in Fig. 18 shows that a nearly linear relation exists between the positive and negative increments in voltage that must be applied to the positive and negative grids, respectively, to maintain a constant operating current, in this case 35 milliamperes.

In Fig. 19, a more extended family of characteristics is shown for a positive-net potential of 65 volts. Across these static characteristics a dynamic load characteristic is drawn for a resistance load of 3400 ohms, which matches the plate resistance of the tube at the operating point

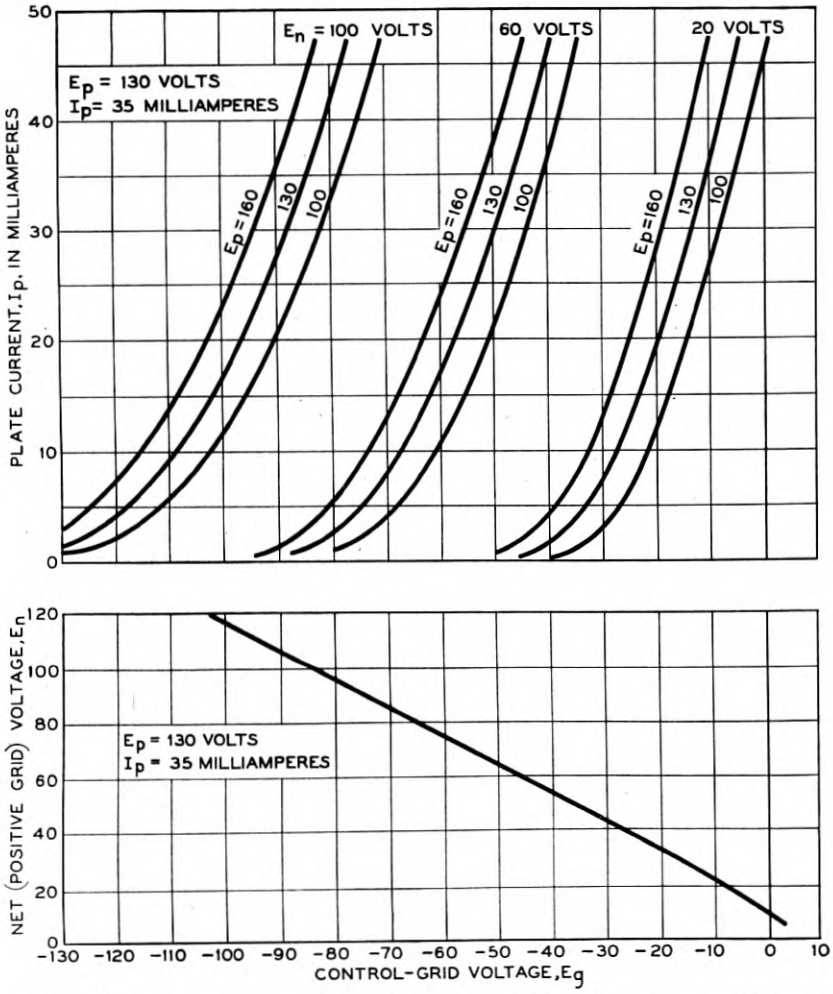


Fig. 18—Plate current-control grid voltage characteristics of a co-planar-grid tube with various values of plate voltage and net (positive-grid) voltage as parameters.

$P$. At this point $E_p = 130$ volts, $E_n = +65$ volts, $E_g = -51$ volts, $I_p = 35$ milliamperes, and $\mu = 5.3$. The turning over of the dynamic characteristic at the top is due to the increasing fraction of the total

space current collected by the positive net. The magnitude of the net current in this region is shown by the family of curves in the lower part of Fig. 19.

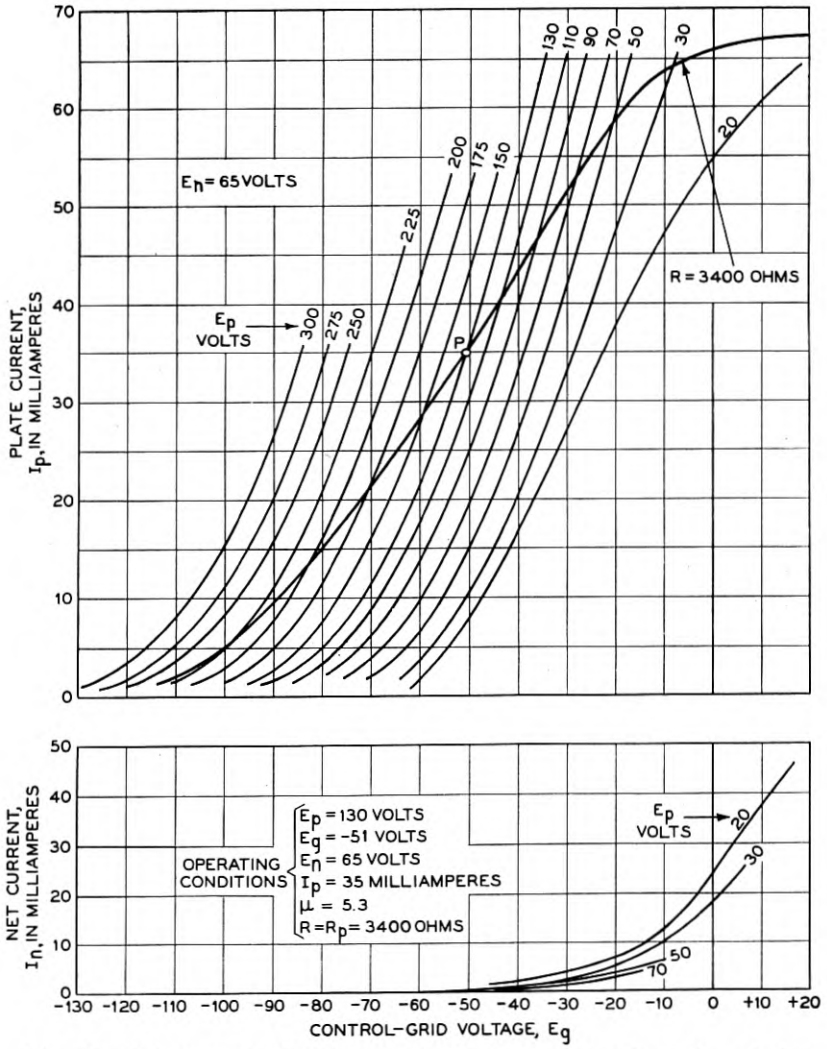


Fig. 19—Plate current-grid voltage and net current-grid voltage characteristics of a co-planar-grid tube. The net (positive-grid) potential is maintained constant at 65 volts. A dynamic characteristic is shown for a resistance load of 3400 ohms.

In Fig. 20, curves are shown giving the power output, and second and third harmonics expressed in decibels below the fundamental, as functions of the peak input when a sinusoidal voltage is applied to the

control grid. The curves were obtained with the indicated values of load resistance, and with the indicated bias voltages applied to the grids. In each case the grid voltages were adjusted to give a plate current of 35 milliamperes. The curves are very similar in form to

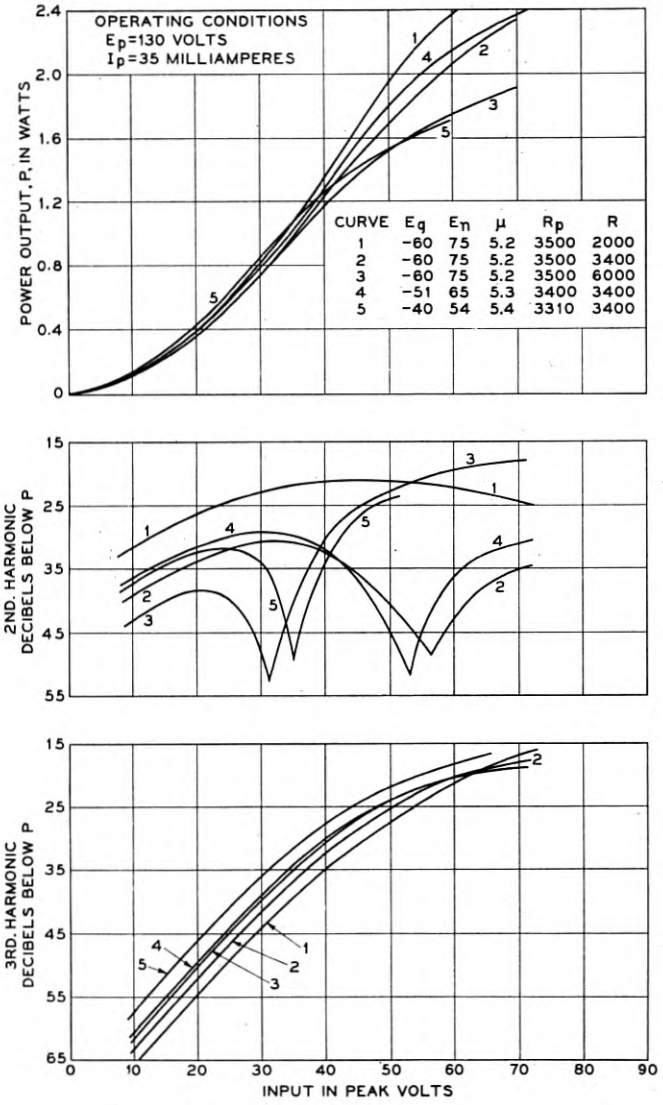


Fig. 20—Co-planar grid tube.

Power output, $P$, in watts, second and third harmonics in db below the fundamental, as functions of peak input when a sinusoidal voltage is applied to the control grid.

those for the power pentode shown in Fig. 12. For the co-planar-grid tube, as for the pentode, the third harmonic is comparatively high and increases with the load resistance, while the second harmonic curves exhibit minima. This is to be expected from the form of the dynamic characteristic which is very similar to that of similar curves for the pentode previously discussed. For example, consider the curves marked 2 in Fig. 20, obtained with $E_p = 130$ volts, $E_n = 75$ volts and

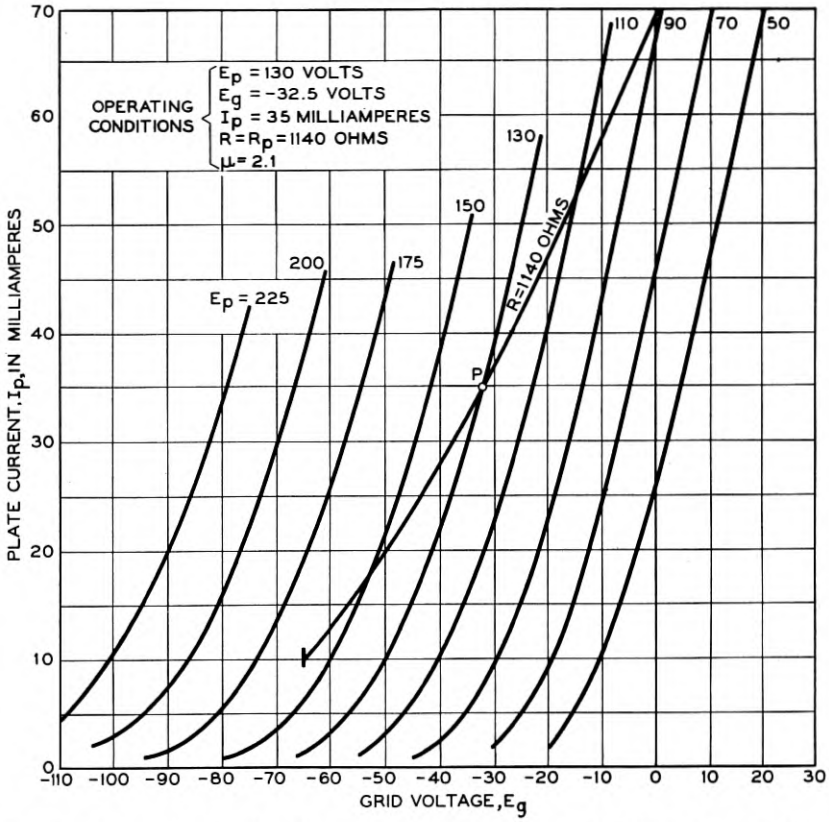


Fig. 21—Plate current-grid voltage characteristics of a triode with a dynamic characteristic shown for a resistance load, $R = R_p = 1140$ ohms.

$E_g = -60$ volts. The output power under these conditions, with an input of 50 peak volts, is 1.68 watts with the second and third harmonics 40 db and 25 db, respectively, below the fundamental. At an input of 30 peak volts, the level of the second harmonic rises to 29 db and that of the third harmonic falls to 41 db below that of the fundamental.

It is of interest to compare the characteristics of this co-planar-grid tube with those of a three-electrode tube designed also to operate at a plate potential of 130 volts and at a plate current of 35 milliamperes. The plate current-grid voltage characteristics of such a triode are shown in Fig. 21. The dynamic characteristic is for a load resistance of 1140 ohms which matches the plate resistance of the tube at the operating point *P*. At zero grid bias on the dynamic characteristic, the plate potential is 92 volts, a drop of only 38 volts from the operating

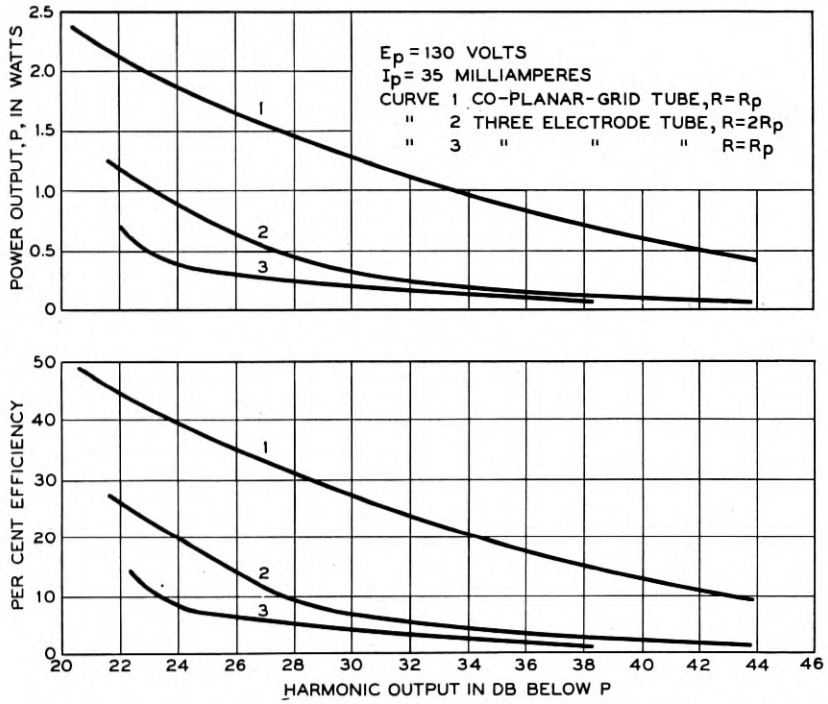


Fig. 22—Curves showing the output power and plate efficiency of a triode and a co-planar-grid tube with varying effective values of second and third harmonics expressed in db below the fundamental.

potential of 130 volts. In the co-planar-grid tube the dynamic characteristic cuts the axis of zero grid bias at a plate potential of 27 volts, a drop of 103 volts from the operating plate potential. The larger plate-voltage variation in the co-planar-grid tube results in higher efficiency than in the triode. The reasons for this difference are identical with those previously discussed in the case of the power pentode.

In Fig. 22, curves are shown comparing the power output and plate-

circuit efficiency of the triode and co-planar-grid tube at varying harmonic levels expressed in db below the fundamental. Curves are shown for the triode with the load resistance, $R$, equal to the plate resistance, $R_p$, and also with $R = 2R_p$. In both tubes the plate voltage is maintained at 130 volts, and the plate current at 35 milliamperes.

With the harmonics 26 db below the fundamental (5 per cent) and with $R = 2R_p$, the output from the triode is 0.65 watt. At the same harmonic level the output from the co-planar-grid tube is 1.65 watts with a resistance load matching its plate resistance. The efficiencies are 14 per cent and 35 per cent respectively.

The most significant differences between the co-planar-grid tube and the power pentode are as follows: (1) The former has a much lower plate resistance which simplifies coupling the tube to its output circuit and is particularly advantageous in applications where it is desirable to operate the tube into an impedance matching its plate resistance. (2) The pentode yields higher amplification and, consequently, requires a correspondingly smaller input voltage to drive it. (3) The plate-circuit efficiency is not greatly different except at low plate voltages where the co-planar-grid tube is somewhat more efficient.

It should be borne in mind in considering the efficiency of output power tubes used at audio frequencies, that the usual criterion of available power is the harmonic content. As has been shown, fundamental physical factors tend to make the harmonics, particularly the odd harmonics, relatively high in pentodes and co-planar-grid tubes. On the other hand, the recent trend in the design of triodes has been toward tubes of very low plate resistance, which operate into comparatively large load resistances. This tends to decrease the harmonics on the one hand and increase the efficiency on the other. The net result has been that triodes of more recent design, particularly those operating at 250 volts or higher, are not far below the pentode or co-planar-grid tube in efficiency.

# Limits to Amplification *

By J. B. JOHNSON and F. B. LLEWELLYN

The amplification obtainable in a vacuum tube amplifier is limited by the noise in the circuit. Of the various sources of noise the most fundamental and inevitable is thermal agitation of electricity. Other sources are the influence of ions and of shot effect and flicker effect on the current in vacuum tubes, poor contacts, mechanical vibration, and hum from a.-c. cathode heating. These noises and their effects in limiting amplification are discussed in this paper. Although the natural noise level of an amplifier is exceedingly low, modern amplifiers have reached such a stage of perfection that their noise levels often are practically at the natural limit.

NOISE in amplifiers is now a familiar term. Any one who has had his favorite radio hour ruined by static knows the effect of an incoherent background of noise. Although static was one of the first noises observed in radio amplifiers, its origin is really outside the circuits. At one time it seemed that other sources of noise of purely local origin, such as poor batteries, loose contacts, gassy tubes and induction from power lines, might be eliminated entirely so that the circuits would be capable of amplifying any signal, no matter how small. It was found, however, that the noise level cannot be lowered indefinitely; that there are limits below which, in the nature of things, noise cannot be reduced.

Of the sources of noise, the most fundamental and inevitable is thermal agitation of electricity. In a perfect amplifier all other noises would be reduced to a level below that of thermal agitation. Next in order comes the influence of ions and of shot effect and flicker effect on the current in the vacuum tubes. Under control to a greater extent, but nevertheless of a malignant nature, are the effects of poor contacts, mechanical vibration, and hum from a.-c. cathode heating. In dealing with these disturbances, the circuit and vacuum tube of the first stage of the amplifier are the most important, for here the signal being amplified is at its lowest level. When the signal is so faint that it is masked by the noise remaining as the natural limit of the circuit, then the only possible remedy is to raise the signal level.

The natural noise level is exceedingly low, yet modern amplifiers have reached such a stage of perfection that their noise levels often are practically at the natural limit. This is true not only of special amplifiers built for experimental purposes, but of many amplifiers

* Published in November 1934 issue of *Electrical Engineering*. Scheduled for presentation at Winter Convention of *A. I. E. E.*, New York City, January 22–25, 1935

used in commercial circuits. The natural limits to amplification which will be discussed in this review are therefore of very practical interest.

## Thermal Agitation[1, 2, 3, 4, 5, 6]

The free charge of any conductor is in random motion in equilibrium with the thermal motion of the molecules of the conductor and this flow of charge creates a random voltage across the terminals of the conductor. This voltage usually is observed in a system composed of an amplifier with an input circuit and an output device. Its mean-square value across the output device is given by the expression

$$V_{to}^2 = 4kT \int_0^\infty RG^2 df, \tag{1}$$

where the symbols have the following meanings:

$k$ is the Boltzmann gas constant and is equal to $1.37 \times 10^{-23}$ watt-second per degree,
$T$ represents absolute temperature, degrees Kelvin,
$R$ represents the resistive component in ohms of the input impedance as measured across the input terminals of the amplifier,
$G$ represents the voltage gain of the amplifier, and is equal to the ratio of voltage across the output device to voltage across the input terminals of the amplifier,
$f$ represents frequency in cycles per second,
$R$ and $G$ are in general functions of frequency.

In the simple case where the amplifier has a constant gain over a frequency range $F$ and no gain outside of this range, and where $R$ is also constant over the same frequency range and is at the normal temperature of 300 degrees, the mean-square noise voltage across the input terminals of the amplifier is

$$V_t^2 = 1.64 \times 10^{-20} RF. \tag{2}$$

This is the voltage that would be produced by a generator supplying to the resistance $R$ the power

$$W = \frac{V_t^2}{R} = 1.64 \times 10^{-20} F. \tag{3}$$

The power $W$, sometimes expressed as $1.64 \times 10^{-20}$ watts per cycle, is independent of $R$ and may be regarded as the apparent input power of the thermal agitation. It depends only on the frequency range of the amplifier, since the temperature cannot be varied conveniently or

[1] For all numbered references see bibliography at end of paper.

very effectively and it sets a lower limit to the possibility of amplifying electrical impulses of any kind. Any signal much smaller than the thermal noise would be masked hopelessly. The only factor under control in the noise equation is the frequency range $F$, which should be no greater than is needed for the transmission of the signal.

An example will illustrate the magnitude involved in this limit to amplification. When the signal is speech requiring a frequency band of 6,000 cycles, then the apparent power generated at the input of the amplifier by thermal agitation is $0.985 \times 10^{-16}$ watts, which is about 138 db below the common reference level of 0.006 watts. (The level of $10^{-16}$ instead of 0.006 watts is being considered as a reference point for the decibel scale in communication circuits. This is approximately the level of thermal noise in a 6,000-cycle channel.) If the input resistance were one megohm the corresponding r.m.s. noise voltage would be 9.94 $\mu$v.

A signal represents a certain amount of available power, and when this is so small that it is near the thermal noise level it must be used efficiently to produce voltage at the grid of the amplifier tube.[7, 8] Let the signal be supplied by a generator of voltage $E$ and internal resistance $R_1$ which delivers power to a load resistance $R_2$, the combination forming the input circuit of the amplifier as shown in Fig. 1. The mean-square signal voltage on the grid of the amplifier tube is

$$V_{sg}^2 = \frac{E^2}{R_1^2}\left(\frac{R_1R_2}{R_1 + R_2}\right)^2. \tag{4}$$

However, the resistance required by equation 2 for the noise is the

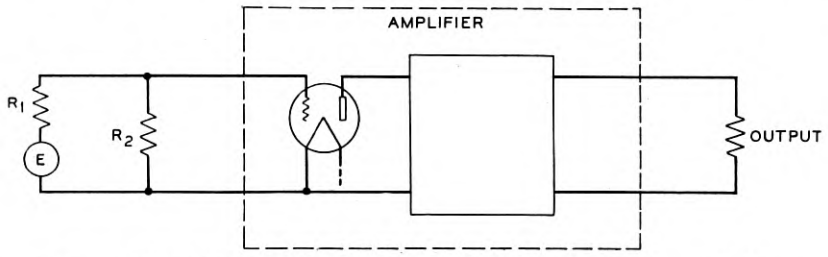


Fig. 1—Schematic diagram of a vacuum tube amplifier showing equivalent input circuit.

combination of $R_1$ and $R_2$ in parallel, so that the mean-square noise voltage on the grid of the amplifier tube is, from equations 2 and 3

$$V_n^2 = WR = W\left(\frac{R_1R_2}{R_1 + R_2}\right). \tag{5}$$

Hence the signal-to-noise ratio is

$$\frac{V_{so}^2}{V_n^2} = \frac{E^2}{WR_1}\left(\frac{R_2}{R_1 + R_2}\right). \tag{6}$$

In general, the internal resistance $R_1$ of the signal generator is fixed, so that $R_2$ is the only available variable. In the usual case of matched impedances where $R_1$ and $R_2$ are equal, the signal-to-noise ratio is 3 db poorer than in the ideal case where $R_2$ is made very large compared with $R_1$. This is one of the few examples where a mismatch of impedances is advantageous. The use of an ideal step-up transformer between $R_1$ and $R_2$ in Fig. 1 will be of no avail, so far as the thermal noise is concerned, because its effect in equation 6 will be only to replace $R_2$ by $R_2/N^2$ where $N$ is the turns ratio of the transformer.

In some systems the impedance at the input of the amplifier is unavoidably small. It may be so small that the voltage of the thermal agitation of the input circuit, even when amplified by the first tube, is lower than the noise voltage originating in the output circuit of that tube. Ideally, the noise in the plate circuit also should be caused by thermal agitation only, and the equations for it have been derived.[9, 10] In practice, however, the noise in the plate circuit is found to be considerably greater, for reasons that will be discussed presently.

### Shot Effect and Flicker Effect Without Space Charge

Early in the study of noise arising in vacuum tubes it was shown [1, 11] that under certain conditions a noise is produced that depends on the fact that the electric current is a flow of discrete particles, the electrons, which are emitted from the cathode in a random manner. The random electron emission produces a statistical fluctuation in the current that flows through the tube and coupling impedance. This fluctuation, called the shot effect (German Schroteffekt, in analogy with the random scattering of shot from a shot gun), appears as noise in the output of the amplifier. When the current in the tube is limited by the rate of emission of electrons rather than by the space charge, so that the resistance of the tube is nearly infinite, then the shot effect produces a mean-square voltage across the output impedance of the amplifier given by [12, 9]

$$V_{so}^2 = 2\epsilon i \int_0^\infty Z^2 G^2 df, \tag{7}$$

in which

$\epsilon$ represents the charge on the electron and is equal to $1.59 \times 10^{-19}$ coulomb,

$i$ represents space current in amperes,
$Z$ represents the magnitude in ohms of the coupling impedance,
$G$ represents the voltage gain of the amplifier from $Z$ to the output,
$f$ represents frequency in cycles per second,

For an amplifier having a flat frequency response curve over frequency range $F$ the expression becomes, for the effective shot noise across the impedance $Z$

$$V_s^2 = 31.8 \times 10^{-20} i Z^2 F. \tag{8}$$

The expression holds quite accurately for tubes in which the cathode is made of either clean or thoriated tungsten and for high-vacuum photo-electric tubes, and it has been used in determining the charge on the electron.[13]

When an oxide coated cathode is used, fluctuations of a larger magnitude [14] are superimposed on the true shot effect. These fluctuations are inappreciable above about 10 k.c., but increase rapidly in magnitude toward the lower frequencies. They also increase with the current at a faster rate than the shot effect fluctuations. This disturbance has been ascribed to a state of flux and change in the activating material on the surface of the cathode,[14, 15] and the phenomenon has been called the "flicker effect" (from the analogy of a flickering candle).

There are two practical circuits in which the pure shot effect may set the ultimate noise level. One of these is the circuit in which the grid of an amplifying tube is left "floating" at its equilibrium potential as usually is done in the first stage of amplifiers used for ion counters and other instruments for measuring very small charges.[18, 10] The grid then emits a few electrons and receives positive ions and electrons from the surrounding space. These currents are very small, but are not subject to space charge limitation so far as the grid is concerned. Because the grid impedance is very high, the shot voltage developed by the small grid current may exceed the thermal voltage of the grid impedance. The second circuit is that in which a photo-electric cell works into the amplifier.[19, 20, 21] Vacuum cells generate shot noise of very nearly the theoretical value given by equation (7), while gas-filled cells give even greater noise.

The total noise generated in the output of the vacuum photo-electric cell is the sum of the shot noise and thermal noise across the coupling resistance $R$, as given by equations (8) and (2). The mean square of the signal voltage, however, is $(\Delta i R)^2/2$ where $\Delta i$ is the amplitude of the current variation. The ratio of signal to noise is then

$$V_{sg}^2/(V_t^2 + V_s^2) = 1.59 \times 10^{18} \frac{(\Delta i)^2}{F} \frac{R}{iR + 0.0516}. \tag{9}$$

This equation shows the expected fact that for a given value of $\Delta i$ it is better to keep the direct photo-electric current small (high modulation). It also brings out the curious result that when the direct voltage drop in the coupling resistance is much more than 1/20 volt the noise is largely shot noise and the signal-to-noise ratio is independent of the coupling resistance, while if this voltage is much less than 1/20 volt the thermal noise predominates and the signal-to-noise ratio is proportional to the coupling resistance.

### Shot Effect and Flicker Effect With Space Charge

When, as in an amplifier tube, the current in the tube is limited partly or wholly by space charge rather than by the cathode temperature, then the conditions are changed [13, 14, 16] in two respects. First, while the electrons still are emitted from the cathode at random times, they must arrive at the plate in a more orderly manner. Simple statistical laws no longer apply, the flow of current is smoother and the fluctuations are greatly reduced. Second, the impedance of the tube is no longer infinite, but has a finite value. The equation for the shot effect (equation (7)) now must be modified,[9, 16, 17] by substituting for the current $i$ the quantity $j(\partial i/\partial j)^2$, where $j$ is the total current emitted by the filament and $\partial i/\partial j$ is the rate of change of space current with emission current for the particular conditions used in the observation of the fluctuating voltage. Furthermore, in place of the coupling impedance $Z$ the effective impedance $Z_e$ of this in parallel with the tube resistance $r_p$ must be used. The equation now reads

$$V^2 = 2\epsilon j\left(\frac{\partial i}{\partial j}\right)^2 \int_0^\infty Z_e^2 G^2 df. \tag{10}$$

In the absence of space charge $j$ and $i$ are identical, $\partial i/\partial j$ is unity, and $Z_e$ becomes $Z$, so that the equation then represents the pure shot effect. With increasing space charge the value of $\partial i/\partial j$ approaches zero and $Z_e$ becomes smaller so that the shot effect becomes very small. Similarly the flicker effect, being connected with the process of emission and not with the subsequent history of the electrons, also is made ineffective by the space charge. In fact, in well designed tubes the fluctuation noise of both shot effect and flicker effect in the space current appears to be reduced to such an extent as to be negligible.

### Ions in the Space Charge

The effect of ions in the grid current already has been discussed. Ions also may cause fluctuations in the plate current of the tube.

The space charge which limits the current between cathode and anode consists of electrons in rapid progress toward the anode. A massive ion placed in this region travels much more slowly and contributes to the space charge for a much longer time than does an electron. While its own charge contributes little to the current, one ion may cause the current to change by the amount of hundreds of electrons during its flight through the space charge region, and the action of many ions would be additive.

Probably most of the ions existing in a tube are positive. Some of them are molecules of residual gas that have lost an electron by collision with an electron of the space current. Residual gas has been found to increase the noise of tubes, especially at the higher pressures. Observations at very low pressures are not conclusive, and it is not certain whether in any modern tubes the noise level is determined by the presence of residual gas.[9, 14, 19, 20, 21, 22, 23, 24]

Positive ions may be emitted also by the cathode. These never can attain a high velocity because they remain in a region of low field intensity. They may be trapped for a time in the region of the potential minimum near the cathode before they finally pass to the grid or possibly become neutralized by an electron. In modern tubes with low temperature filaments the effect of these ions is reduced greatly, yet still may account for a large part of the difference between the observed tube noise and the theoretical thermal noise of tubes.[14, 25, 26, 27, 28]

## Noise in Commercial Tubes

Noise generated in an amplifier should consist largely of the thermal noise of the input circuit, to which is added the noise produced in the plate circuit of the first tube. It is convenient to consider that the tube noise comes, not from the plate circuit of the tube, but from a fictitious resistance $R_G$ in series with the resistance $R_e$ of the input circuit.[2] The effective thermal noise of the input circuit then is given by the expression

$$V^2 = 4kTF(R_e + R_G) = 1.64 \times 10^{-20}F(R_e + R_G). \qquad (11)$$

The tubes therefore may be rated conveniently in terms of $R_G$. The transformation to volts or to watts can be accomplished readily by equations (2) or (3). If, with a given tube and circuit, $R_G$ approaches or exceeds $R_e$ in value, the tube is responsible for an appreciable part of the total noise. The choice of another tube in which the ratio of $R_e$ to $R_G$ is more favorable then may be considered.

For the calculation of tube noise several formulas have been proposed, either entirely empirical [2] or with some basis on theory.[9, 24, 29] These formulas generally fail in the prediction of noise in tubes for the reason that the greater part of the noise in practical tubes is caused by things that have not been included in theory and that are still in a state of flux so far as manufacturing is concerned. It is best, therefore, to rely only on actual measurements of the noise in specific types of tubes. With modern tubes, the noise level of a given type of tube can be represented reasonably well by measurements made on a small number of samples.

Published data on noise of tubes are rather meager. The best series of measurements is that of Pearson,[10] which covers four Western Electric tubes at different frequency bands. These tubes are known commercially as types 102*G*, 262*A*, 264*B*, and 259*B*. The General Electric tube type *PJ*-11, designed specially for work at low frequencies, was studied by Metcalf and Dickinson.[24] They also give data, for the low frequency region, on the tubes known commercially as types 222, 240, 201, and 112. Johnson and Neitzert [30] have given data for the *PJ*-11 and the type '38 tube. Certain British tubes were studied by Moullin and Ellis,[29] and of these the type *AC*/2*HL* tube was found to have the lowest noise level. Brintzinger and Viehmann [31] studied a few German tubes. Of these the type *RE*-084 appears to have the lowest noise rating, but the data cannot be reduced to absolute measure.

In many of these studies the tubes were operated at voltages different from those usually employed. For these, the original papers should be consulted. In general for the best triodes $R_G$ has a value of a few thousand ohms, while for screen tubes it has a value of a few tens of thousands. At the lowest voice frequencies the values may be somewhat greater.

## Other Sources of Noise

While the more fundamental sources of noise have been discussed, it may be well to add some remarks on a few types of disturbance that often can be eliminated.

### *Noise From A-C Cathode Heating* [32–39]

The indirectly heated cathode may be operated on alternating current when the tube is employed in radio frequency circuits. In audio amplifiers with gains in excess of 50 db, additional precautions must be taken to reduce the effects of the electric and magnetic fields of the heater and of coupling impedance between the heater and the other electrodes. Even under the best conditions, however, the hum

level is of the order of 20 db above the tube noise measured with d-c heating.

### *Noise From Vibration* [32, 33, 39, 40]

Mechanical vibration changes the relative positions of the tube elements and hence causes disturbing noise. This is especially objectionable at audio frequency, although a radio frequency carrier may become modulated sufficiently to produce noise.

The remedy, used in the so-called "low microphonic" tube designs, is to stiffen the construction of the tube elements and to apply damping to their vibration, as well as to cushion the tube by a suitable mounting and to shield it from sound waves. The indirectly heated cathode is superior to the filamentary cathode in regard to noise from vibration.

### *Noise From Poor Insulation* [32, 33, 39]

Noise arises from resistance changes at contacts and across thin films of conducting material deposited on insulating supports in the vacuum tube. Leaky capacitors may produce a similar noise.

### *Noise From Faulty Resistances*

Many resistors in which the resistance element is a thin film are sources of noise. If no current flows in them only thermal noise is generated, but when direct current passes through them more noise is produced. The noise voltage is roughly proportional to the direct current. These resistors must be chosen carefully for circuit branches where current flows.

### Signal-to-Noise Ratio [7, 29, 41, 42]

In so far as noise is concerned, the merit of a transmission system is dependent not only on the amount of noise present, but also on the strength of the signal, so that a determination of the ratio of the signal level to the noise level is necessary. Fortunately, this ratio has a reference value for any given transmission system determined uniquely by the ratio of the signal to thermal noise in the input circuit. The ratio of the signal level to noise level is here the greatest the ratio ever can attain, because noise that originates at subsequent points in the amplifier contributes to the noise level without increasing the signal.

This fact provides a basis for the rating of amplifiers, the thermal noise of the input circuit being used as a comparison signal. For example, the noise output of an amplifying system may be 0.3 mw which falls to 0.2 mw when the input circuit is short-circuited. The thermal noise from the input circuit is then the difference between

these two readings, namely, 0.1 mw, and the signal-to-noise ratio of the actual system is 3 times, or 4.8 db worse than its best possible value with a given signal. These data may be expressed in terms of an equivalent input resistance which has the advantage that the amplifying properties of the tube have been taken into account. This leaves for the engineer only the problem of selecting a tube having an input capacity of such a value that the construction of a relatively high impedance input circuit is possible.

So far, the discussion has been based upon the properties of amplifiers only, no mention being made of the effects of modulators, detectors, frequency converters, and other nonlinear devices on the signal-to-noise ratio. A detailed discussion of the noise in such devices is beyond the scope of this paper, but the relations in the most commonly used ones may be indicated and their general properties outlined.

First, consider a system composed of a radio frequency amplifier followed by a detector and a pair of headphones. A certain amount of noise will be heard in the phones if the gain of the amplifier is great enough. This noise is caused by the various components of the radio frequency noise beating together in the detector to form audio frequency components. Next, suppose that an unmodulated carrier is introduced into the amplifier. It will be observed that the audio noise in the phones increases. The increase in audio noise is produced by the radio frequency carrier beating with the radio frequency noise components and this increase is proportional to the strength of the carrier.

If a small percentage of modulation is added to the carrier, the audio signal-to-noise ratio in the phones will be determined by the properties of the amplifier in the same way as though the system were a straight amplifier without any detector. Comparison of the actual system with the ideal may be made by introducing the carrier into one of the amplifier stages subsequent to the input, and then measuring the audio noise with the input circuit in its normal condition and again with the input circuit short-circuited. The ratio of these two energy values gives the ratio of the equivalent input resistance of the actual system to the equivalent input resistance of the noisy amplifier alone. The ratio of the ideal signal-to-noise ratio to the actual one may be found by dividing the difference between the two audio energy readings by the reading taken with the input circuit in its normal condition.

If the percentage of modulation of the carrier is large, the system will be noisier because there will be appreciable audio noise components caused by beats between the side bands and the radio frequency noise components. Again, if the carrier level is not large compared with

the noise level in the amplifier, the system will be noisier because the beats between the noise components are appreciable compared with the beats between carrier and the noise.

The same considerations apply to the first detector in a double-detection receiving system. If, as is usual, the beating oscillator voltage is large compared with the noise components, then the frequency band of the noise will be shifted in position in the same manner as the signal, and the signal-to-noise ratio of the system will be unchanged by the frequency conversion.

The signal-to-noise properties of any system are considered satisfactory when the total output noise differs only slightly from that produced by thermal agitation in the input circuit alone, and this difference may be measured by eliminating the input thermal noise (as by the short-circuit method) and noting the change produced in the output noise.

BIBLIOGRAPHY

1. W. Schottky, "Ueber spontane Stromswankungen in vershiedenen Elektrizitätsleitern" ("On Spontaneous Current Fluctuations in Various Electric Conductors"), *Ann. d. Phys.*, v. 57, 1918, p. 541–67.
2. J. B. Johnson, "Thermal Agitation of Electricity in Conductors," *Phys. Rev.*, v. 32, 1928, p. 97–109.
3. H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," *Phys. Rev.*, v. 32, 1928, p. 110-13.
4. E. K. Sandemann and L. H. Bedford, "The E.M.F. of Thermal Agitation," *Philosophical Mag.*, v. 7, 1929, p. 774–82.
5. N. H. Williams and D. A. Wilbur, "Thermal Agitation of Electrons in a Metallic Conductor," *Science*, v. 76, 1932, p. 519–20.
6. E. B. Moullin and H. D. N. Ellis, "A Measurement of Boltzmann's Constant by Means of the Fluctuations of Electron Pressure in a Conductor," *Cambridge Philosophical Soc. Proc.*, v. 28, 1932, p. 386–402.
7. F. B. Llewellyn, "A Rapid Method of Estimating the Signal-to-Noise Ratio of a High Gain Receiver," *I. R. E. Proc.*, v. 19, 1931, p. 416–20.
8. H. S. Black, "Stabilized Feed-Back Amplifiers," *Elec. Engg.*, v. 53, 1934, p. 114–20.
9. F. B. Llewellyn, "A Study of Noise in Vacuum Tubes and Attached Circuits," *I. R. E. Proc.*, v. 18, 1930, p. 243–65.
10. G. L. Pearson, "Fluctuation Noise in Vacuum Tubes," *Physics*, v. 5, 1934, p. 233–43 and *Bell Syst. Tech. Jour.*, v. 13, 1934, p. 634–53.
11. W. Schottky, "Zur Berechnung und Beurteilung des Schroteffektes" ("On the Calculation and Critical Examination of the Shot Effect"), *Ann. d. Phys.*, v. 68, 1922, p. 157–76.
12. T. C. Fry, "The Theory of the Schroteffekt," *Franklin Inst. Jour.*, v. 199, 1925, p. 203–20.
13. A. W. Hull and N. H. Williams, "Determination of Elementary Charge E From Measurements of Shot Effect," *Phys. Rev.*, v. 25, 1925, p. 147–53.
14. J. B. Johnson, "The Schottky Effect in Low Frequency Circuits," *Phys. Rev.*, v. 26, 1925, p. 71–85.
15. W. Shottky, "Small Shot Effect and Flicker Effect," *Phys. Rev.*, v. 28, 1926, p. 75–103.
16. E. W. Thatcher and N. H. Williams, "Shot Effect in Space Charge Limited Currents," *Phys. Rev.*, v. 40, 1932, p. 474–96.
17. E. W. Thatcher, "On the Reduction of Shot Effect Fluctuations," *Phys. Rev.*, v. 40, 1932, p. 114–15.
18. R. L. Hafstad, "The Application of the FP-54 Pliotron to Atomic Disintegration Studies," *Phys. Rev.*, v. 44, 1933, p. 201–13.

19. B. A. Kingsbury, "The Shot Effect in Photoelectric Currents," *Phys. Rev.*, v. 38, 1931, p. 1458–76.
20. F. V. Orbán, "Schroteffekt und Wärmegräusch im Photozellenverstärker" ("Shot Effect and Thermal Noise in Photoelectric Cell Amplifiers"), *Zeits. f. Techn. Phys.*, v. 13, 1932, p. 420–4 and v. 14, 1933, p. 137–43.
21. E. Steinke, "Natürliche Schwankung schwächster Photoströme" ("Natural Fluctuations of Very Small Photoelectric Currents"), *Zeits. f. Phys.*, v. 38, 1926, p. 378–403.
22. S. Ballantine, "Fluctuation Noise Due to Collision Ionization in Electronic Amplifier Tubes," *Phys.*, v. 4, 1933, p. 294–306.
23. N. P. Case, "Receiver Design for Minimum Fluctuation Noise," *I. R. E. Proc.*, v. 19, 1931, p. 963–70.
24. G. F. Metcalf and T. M. Dickinson, "A New Low Noise Vacuum Tube," *Phys.*, v. 3, 1932, p. 11–17.
25. J. S. Donal, "Abnormal Shot Effect of Ions of Tungstous and Tungstic Oxide," *Phys. Rev.*, v. 36, 1930, p. 1172–89.
26. H. N. Kozanowski and N. H. Williams, "Shot Effect of the Emission From Oxide Cathodes," *Phys. Rev.*, v. 36, 1930, p. 1314–29.
27. L. P. Smith, "Effect of Positive Ion Shot Effect on Space Charge Limited Currents," *Phys. Rev.*, v. 35, 1930, p. 1430.
28. E. W. Thatcher and N. H. Williams, "Shot Effect in Space Charge Limited Currents," *Phys. Rev.*, v. 39, 1932, p. 474–96.
29. E. B. Moullin and H. D. M. Ellis, "The Spontaneous Background Noise in Amplifiers Due to Thermal Agitation and Shot Effects," *Jour. I. E. E.*, v. 74, 1934, p. 323–356.
30. E. A. Johnson and C. Neitzert, "The Measurement of Small Alternating Voltages at Audio-frequencies," *Rev. Sci. Inst.*, v. 5, 1934, p. 196–200.
31. W. Brentzinger and H. Viehmann, "Das Rauschen von Empfängern" ("Noise in Radio Receivers"), *Hochfr. und Elektroauk.*, v. 39, 1932, p. 199–207.
32. M. J. Kelly, "Vacuum Tube and Photoelectric Tube Developments for Sound Picture Systems," *Jour. Soc. Motion Picture Engrs.*, v. 18, 1932, p. 761–81.
33. J. O. McNally, "Analysis and Reduction of Output Disturbances Resulting From the Alternating-Current Operation of the Heaters of Indirectly Heated Cathode Triodes," *I. R. E. Proc.*, v. 20, Aug. 1932, p. 1263–83.
34. H. M. Freeman, "A Practical Alternating-Current Radio Receiving Tube," *Elec. Jour.*, v. 19, Dec. 1922, p. 501–5.
35. F. G. McCullough, "Thermionic Tubes," *I. R. E. Proc.*, v. 10, Dec. 1922, p. 468–85.
36. A. W. Hull, "A Combined Kenotron Rectifier and Pliotron Receiver Capable of Operation by Alternating Current Power," *I. R. E. Proc.*, v. 11, April 1923, p. 89–96.
37. B. F. Meissner, "AC as a Filament Supply Source," *Radio Broadcast*, v. 10, Feb. 1927, p. 393–6, March 1927, p. 495–7.
38. W. J. Kimmell, "The Cause and Prevention of Hum in Receiving Tubes Employing AC Direct on the Filament," *I. R. E. Proc.*, v. 16, Aug. 1928, p. 1089–1106.
39. D. B. Penick, "The Measurement and Reduction of Microphonic Noise in Vacuum Tubes," *Bell Sys. Tech. Jour.*, v. 13, Oct. 1934, p. 614–33.
40. A. C. Rockwood and W. R. Ferris, "Microphonic Improvement in Vacuum Tubes," *I. R. E. Proc.*, v. 17, Sept. 1929, p. 1621–32.
41. S. Ballantine, "Fluctuation Noise in Radio Receivers," *I. R. E. Proc.*, v. 18, Aug. 1930, p. 1377–87.
42. E. N. Dingley, "A Common Source of Error in Measurements of Receiver Selectivity," *I. R. E. Proc.*, v. 22, May 1934, p. 546.

# Vacuum Tubes as High-Frequency Oscillators *

By M. J. KELLY and A. L. SAMUEL

Vacuum tubes as oscillators and amplifiers at frequencies greater than 100 megacycles (3 m.) are considered in this paper. The type of construction used in a large number of different tubes, and the characteristics of the tubes, are presented. Circuits for operating the tubes also are considered and the theory of operation and the factors limiting ultra-high frequencies are discussed. Principal attention is given to the tubes as oscillators, with brief consideration of the problem of amplification.

THE three types of oscillation generators which at present are the most efficient in the range from 100 megacycles to 3000 megacycles per second will be discussed in the following survey. These are: the negative grid tube which at lower frequencies is the conventional regenerative oscillator, the positive grid or Barkhausen oscillator, and the "magnetron" oscillator. The amplification problem will be briefly discussed. Because of the present unsettled state of the theory, only the most elementary and generally accepted part will be included. Much theoretical and experimental work remains to be done before knowledge of the mechanism of oscillation and amplification in this frequency range will be satisfactory. As is often the case, the empirical knowledge of some of these mechanisms has outdistanced the theoretical interpretation.

## The Negative Grid Oscillator

The conventional thermionic vacuum triode, whether it be a large water-cooled power tube or a small receiving tube, may be used as a generator of oscillations varying in frequency from a few cycles per second to some 20 or 30 megacycles with substantially undiminished efficiency and output. In this range the frequency at which a tube is to be employed is a factor of almost negligible importance in the determination of its characteristics and its form. Beyond this range, however, frequency plays an increasingly important part, and as one approaches 300 megacycles, it becomes the most important factor in the determination of tube design.

When an attempt is made to operate a standard triode at increasingly high frequencies it is found that the output and efficiency begin to decrease. The frequency at which this is first observed will depend

* Published in November 1934 issue of *Electrical Engineering*. Presented at Winter Convention of *A. I. E. E.*, New York City, January 22–25, 1935.

upon the design of the tube but it will usually be in the 10- to 60-megacycle range. By successive modifications of the circuit arrangement and size this decrease in power output and efficiency can be minimized. With optimum circuit arrangements, however, this decrease continues until finally a frequency is reached beyond which oscillations can no longer be produced.

In Fig. 1 are shown typical data for a standard 75-watt tube, the Western Electric type 242*A*, operated with reduced potentials over

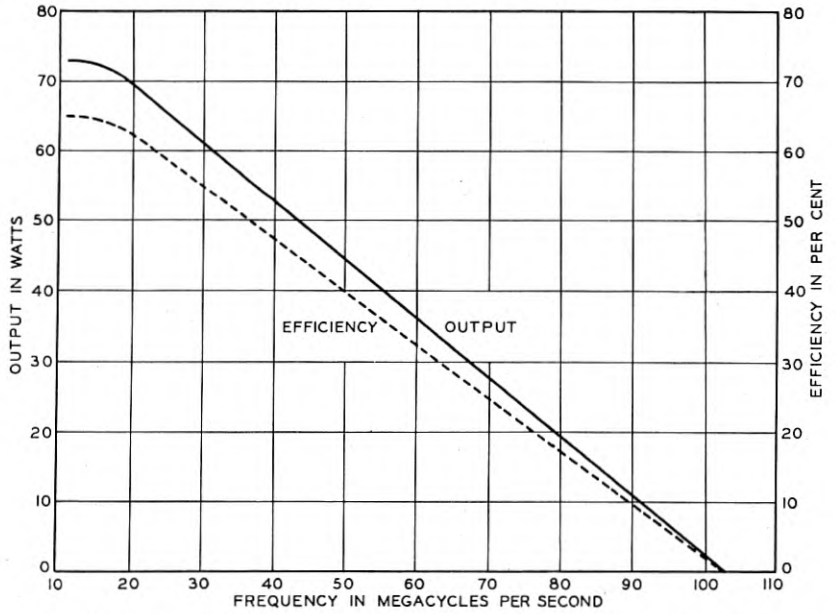


Fig. 1—Power output and anode efficiency as a function of frequency for a standard triode. These curves are typical of all tubes as they approach their upper limiting frequency.

the frequency range from the point where oscillation frequency noticeably affects performance to the point where oscillations can no longer be produced. The plate potential was held constant at 750 volts throughout the entire frequency range. The oscillation circuit was modified at each point in order to obtain maximum output and efficiency, keeping the anode dissipation within the maximum rating of 100 watts and the anode current within the maximum rating of 0.150 amp. It can be seen from the curves that the output and efficiency are independent of the frequency until about 20 megacycles is reached, when they begin to decrease. At 100 megacycles the output power is only 2.5 watts and the efficiency only 2 per cent. The tube will not oscillate at 105 megacycles.

*Effect of Energy Losses on Performance*

A tube operating in the range where frequency affects performance must withstand energy losses, and the resulting heating within its structure, which occur to only a negligible degree at the lower frequencies. Some of these losses are due to dielectric hysteresis in the insulating materials of the tube, particularly in the portions of the glass supporting stem or bulb which lies between the tube leads. The glass is sometimes so softened by the heat thus developed that it is punctured by the outside air pressure. Losses also occur in the auxiliary metallic parts of the tube structure due to the increased eddy currents that occur at high frequencies. Losses in the tube electrodes and their lead-in wires are also greatly increased due to skin effect which increases their resistance, and due to the increased charging current required by the inter-electrode capacitances. The increased lead temperature, depending upon its amount, will cause a more or less rapid deterioration of the lead-to-glass seals which may ultimately destroy the vacuum. In order to protect the tube from damage because of these new types of energy dissipation, the operating potentials and currents must be reduced to values less than those established for low-frequency operation. Some manufacturers are now giving special ratings on such of their standard tubes as may be used at ultra-high frequencies. These ratings should be adhered to when operating in this range.

*Effect of Circuit on Performance*

The decrease in output power and plate efficiency which sets in with the increase in frequency, while due in part to the losses described above, and to the rapid increase in radiation losses, is also due to two additional effects of fundamental importance. The first to become evident, with increasing frequencies, is circuital in nature. This can be explained by reference to the conventional oscillator shown in Fig. 2. The frequency of such an oscillator is given by

$$f = \frac{1}{2\pi\sqrt{LC}},$$

where $L$ and $C$ are the effective inductance and capacitance of the oscillating circuit. In the lower frequency range, the $LC$ product which determines the frequency is substantially equal to $L_0C_0$, that is, to the product of the inductance and capacitance of the external circuit. The inductance of the tube leads and the capacitance between its electrodes (indicated by the dotted lines in the figure) play a negligible rôle. In order to tune the circuit to higher and higher frequencies,

the capacitance $C_0$ is first reduced and finally eliminated, leaving the inter-electrode capacitance as the only capacitance in the oscillating circuit. The tube leads then form a part of the main oscillating circuit, in which large circulating currents must exist for stable operation. For a further increase in frequency the external inductance $L_0$

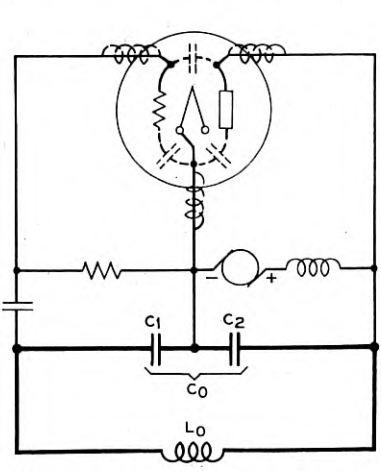


Fig. 2 (left)—A standard Colpitts oscillator circuit. The heavy lines indicate the main oscillating circuit. The dotted portions represent the inter-electrode capacitances and lead inductances which play a minor rôle at low frequencies.

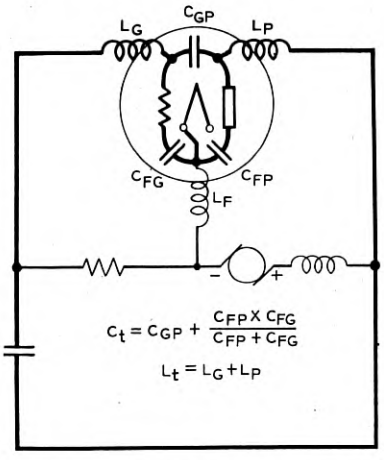


Fig. 3 (right)—The limiting circuit with the external tuning capacitance eliminated and the external inductance reduced to a short circuiting bar between the grid and plate leads. The main oscillating circuit, indicated by the heavy lines, is seen to include the inter-electrode capacitances and lead inductances.

must be reduced and, in the limit, it becomes the shortest possible connection between the grid and plate terminals. The oscillating frequency for this limiting circuit, shown in Fig. 3, is determined by the product of the lead inductance $L_t$ and the inter-electrode capacitance $C_t$; that is,

$$f_0 = \frac{1}{2\pi\sqrt{L_t C_t}},$$

where $L_t$ is the sum of the grid and plate lead inductance and $C_t$ is the total grid-plate capacitance. Even before this frequency limit is reached the output power and plate efficiency are seriously reduced by the lack of full control over the relative amplitude and phase of the alternating grid and plate potentials. Whereas the ratio of these amplitudes is controlled in the circuit shown in Fig. 2 by the condensers $C_1$ and $C_2$, it is determined in the limiting case primarily by

the fixed ratio of the grid-filament to plate-filament inter-electrode capacitance. Most tubes made especially for ultra-high-frequency use are constructed so as to minimize these circuit limitations by a reduction in the inter-electrode capacitance and lead inductance and by adjusting the capacitance ratio.

*Effect of Transit Time on Performance*

The second fundamental effect has to do with the time required for the electrons to travel from the cathode to the anode within the tube structure. This time, the so-called transit time, is very small in present-day commercial types of power tubes, usually much less than one micro-second. Obviously at low frequencies it can be neglected and, in fact, for many tubes it still plays a minor rôle either in determining the output and efficiency in the high-frequency range or in establishing the limiting frequency for oscillations. When the frequency range of oscillation of a tube is extended by an adequate decrease in energy losses and by improvements in electrical design, transit time becomes a dominating factor in the reduction of output power and efficiency and in establishing the limiting frequency of oscillation.

This comes about in two ways. In the first place, the relative phase of the alternating grid and plate potentials for best operation must be altered to compensate for the time required for the electrons to travel from the region in which the grid has its greatest effect upon their motion to the region in which their motion has the greatest effect upon the plate current. The available control over these phases is usually insufficient to permit a realization of the optimum adjustment. In terms of the measured characteristics of the tube, the transconductance has become complex. But even with the optimum phase adjustment the efficiency is reduced by losses which occur because of the variations in grid and plate potentials during the transit time. Electrons arriving at the plate will in general have velocities greater than the velocity corresponding to the potential of the anode at the instant of their arrival. The excess energy corresponding to the greater velocity is obtained from the oscillating circuit and is dissipated at the plate in the form of heat. Again in terms of the measured characteristics, the input conductance has been increased above its low-frequency value.

The mechanism which enables electrons to take energy from the oscillatory circuit in their passage across the tube is evident from a consideration of a somewhat simplified case as shown in Fig. 4. Assume that the anode is held at a constant positive potential of 100

volts, and that the grid is held at 50 volts positive just long enough to allow an electron to come from the cathode to the grid plane (very near one of the wires), where its velocity will correspond to a fall of 50 volts. The potential of the grid is then suddenly changed to 50 volts negative. The electron will then fall through an additional potential difference of 150 volts, arriving at the anode with a velocity corre-

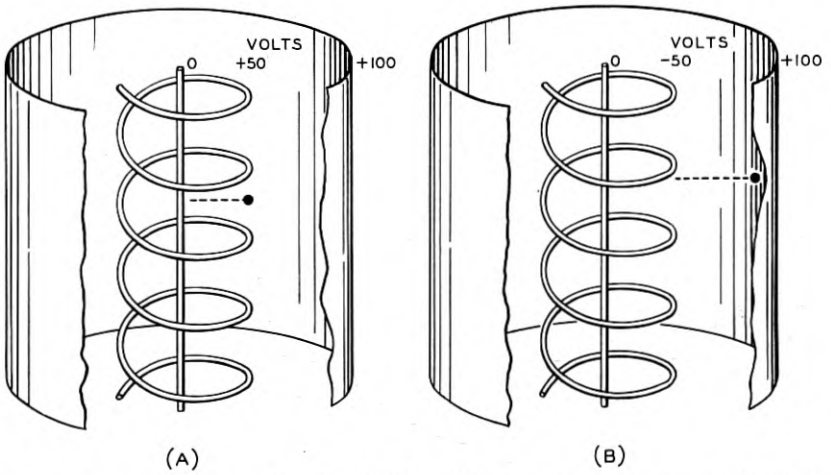


Fig. 4—Illustrating the mechanism which enables electrons to take energy from the oscillating circuit.

sponding to 200 volts, producing just twice as much heat as it would have done had the grid potential not been changed during the transit time. This added energy must come from the source which produced the change in the grid potential. In the actual case the change in grid potential is not abrupt but a similar loss occurs. This limits the useful frequency range of a tube to values for which the oscillation period is long compared to the electron transit time.

*Special Designs Required for Different Ranges of High Frequencies*

Most standard power tubes reach their upper frequency limit of oscillation somewhere in the 10- to 100-megacycle frequency range. For frequencies above this, specially designed tubes are required. The frequency range in which a given design is near the optimum is limited. Therefore, there is a succession of tubes, each rated for a band of frequencies. Characteristics such as a high mutual conductance and a sharp cut-off which make a tube a good oscillator at low frequencies, while still of importance at ultra-high frequencies, are apt to be secondary to the special frequency requirements

Although some progress has been made in the modification of conventially designed water-cooled tubes for use above 100 megacycles, more attention has so far been given to the development of radiation-cooled tubes for this frequency range.

A departure from conventional design with increasing frequency is illustrated by a radiation-cooled tube described by McArthur and Spitzer [1] in which the ratio of the plate diameter to the plate length

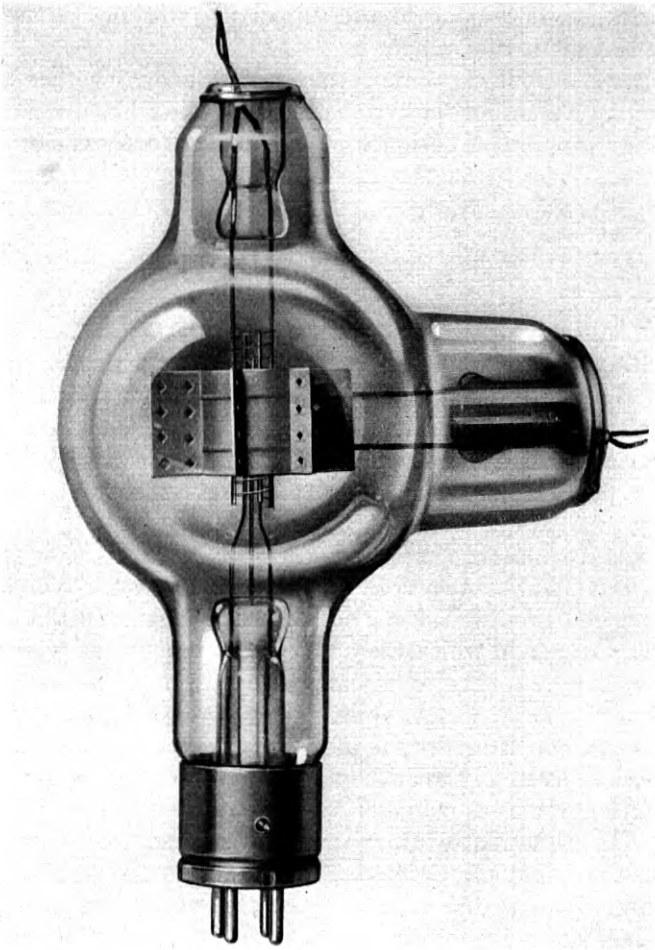

Fig. 5—A radiation cooled tube for use in the frequency range from 60 to 180 megacycles per second. Note the large ratio of the plate diameter to plate length and the special arrangement of the leads.

[1] For all numbered references see list at end of paper.

is much larger than for the conventional tube. Radiating fins are employed to compensate for the decrease in heat radiating ability of the plate which would otherwise occur because of its short length. In Fig. 5 is shown a photograph of this tube. It will be noted that the tube electrodes are supported directly from their leads. The complete absence of auxiliary supporting members either of metal or of insulating material and the large size of leads reduce radiation, eddy current, and conduction current losses. That portion of the inter-electrode capacitances due to the supporting structure is also made small by this method of support.

The inter-electrode capacitances are given below, together with the corresponding values for the type 242*A* tube, which has the same plate dissipation rating but is designed for use at lower frequencies:

| | High Frequency Tube | 242*A* |
|---|---|---|
| Plate to grid | 3 $\mu\mu$f | 13.0 $\mu\mu$f |
| Grid to filament | 2 $\mu\mu$f | 6.5 $\mu\mu$f |
| Plate to filament | 1 $\mu\mu$f | 4.0 $\mu\mu$f |

The decrease in capacitance by a factor of approximately 4 makes possible a much greater improvement in performance in the 60- to 150-megacycle-per-second frequency range than the corresponding degradation in performance due to the lower mutual conductance and the increased electron transit time resulting from the increased spacing. The material increase in plate impedance makes it necessary to employ an anode potential approximately twice as great with the high-frequency tube.

Output and efficiency curves are shown in Fig. 6. (For the sake of uniformity, curves taken from published papers have, in most cases, been redrawn.) The particular shape of the output curve is due to the manner in which the applied anode potential was reduced with increasing frequency to minimize the danger of tube failure from the increased energy losses which occur at high frequencies. The limiting frequency as set by the inter-electrode capacitances and lead inductances is given by the authors as 230 megacycles. An extension of the efficiency curve to higher values indicates that the tube will probably fail to oscillate before this limit is reached. From this it can be inferred that the decrease in efficiency in the range from 150 to 200 megacycles is due largely to the effect of the relatively large transit time, since the authors' method of arriving at the output by taking the difference between the measured input and the measured plate losses includes circuit and lead losses as a part of the output. A comparison of the data of Fig. 1 and Fig. 6 shows that at 100 megacycles the output of the type 242*A* tube is 2 watts and the corre-

sponding output of the high-frequency tube is approximately 86 watts with substantially the same plate loss. This strikingly illustrates the improvement obtained by taking into account the factors which become important in the high-frequency range.

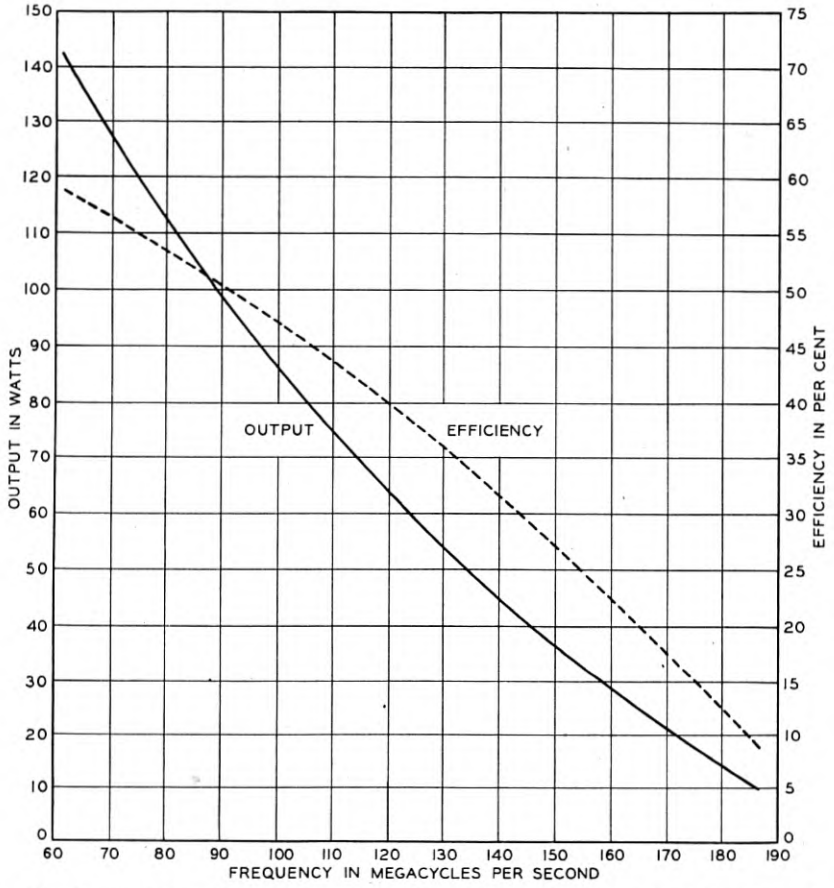


Fig. 6—Output and efficiency as a function of frequency for the tube shown in Fig. 5. A comparison of these curves with those shown in Fig. 1 illustrates the improvement obtained by taking account of those factors which become important at high frequencies.

### *Transit Time Becomes More Important*

In extending the frequency range to 300 megacycles the importance of electron transit time becomes relatively greater. It must be kept as low as possible even at the expense of relatively higher inter-electrode capacitances. In the tube just described for the frequency range around 100 megacycles the reverse procedure is followed; in

order to make the inter-electrode capacitances as small as possible, transit times are increased. Fay and Samuel [2] in a recent paper presented before the International Scientific Radio Union (U.R.S.I.) describe a tube designed for use at 300 megacycles which well illustrates this point. The tube is shown in Fig. 7. It differs from the one

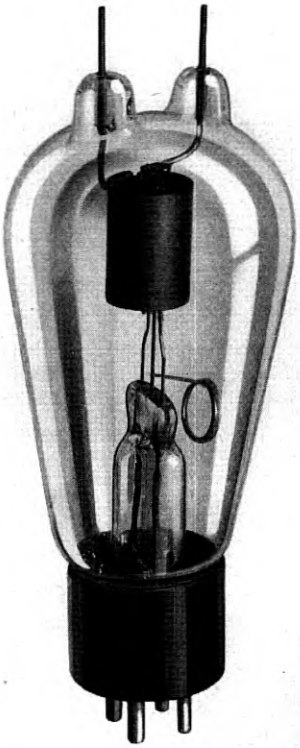

Fig. 7—A tube for use at frequencies up to 350 megacycles.

previously discussed in the close spacings between elements, particularly between the grid and filament. The lead length is further decreased and lead diameter made considerably larger in order to decrease lead inductance and resistance. The inter-electrode capacitances are:

| | |
|---|---|
| Plate to grid | 2.5 $\mu\mu$f |
| Grid to filament | 2.0 $\mu\mu$f |
| Plate to filament | 0.67 $\mu\mu$f |

While these capacitances are substantially the same as for the tube shown in Fig. 5, the limiting frequency, as set by circuit resonance, is

somewhat beyond 400 megacycles as contrasted with 230 megacycles for the other tube. This is due, primarily, to the material decrease in lead inductance. The decreased losses resulting from the minimized transit time more than compensate for the increased circuit loss resulting from the required higher charging currents to the inter-electrode capacitances. Its mutual conductance of 2200 micro-

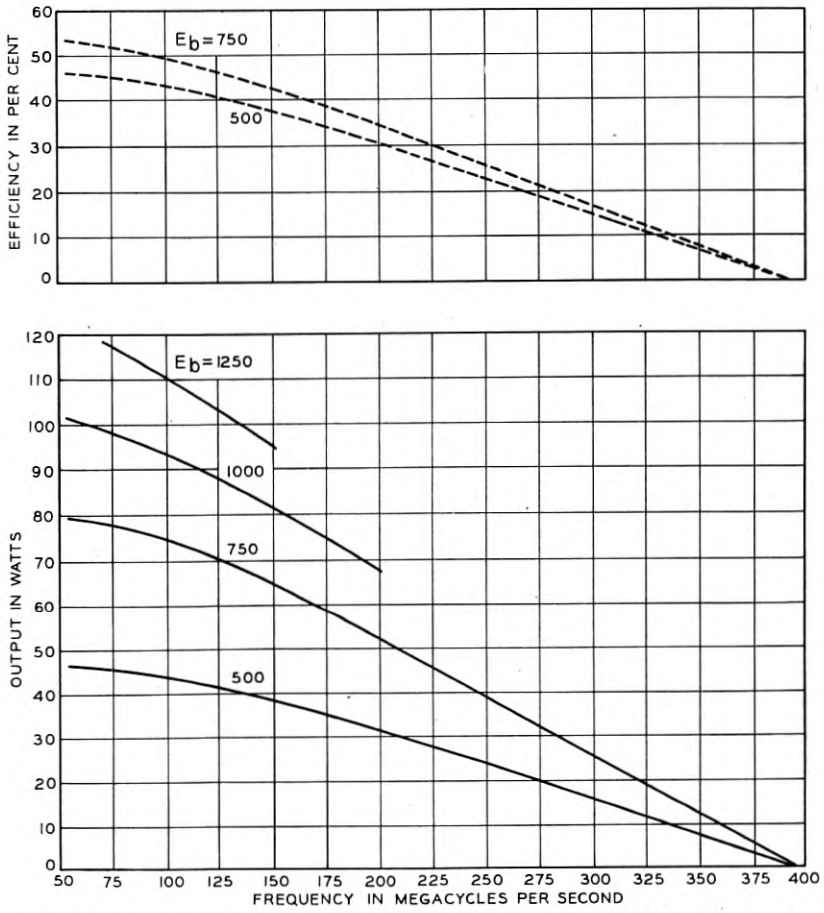


Fig. 8—Output and efficiency curves for 2 tubes of the type shown in Fig. 7.

ohms and the sharp cut-off shown by the static characteristics indicate that those electrical characteristics which are important for efficient oscillators in the low-frequency range have not been sacrificed in meeting the requirements of 300-megacycle operation.

Output and efficiency curves for this tube are shown in Fig. 8.

These data are for two tubes operated in the push-pull circuit shown in Fig. 9. The output shown represents only useful power, since it is the photometrically measured power consumed in a lamp load. It will be noted that, whereas the maximum output at 100 megacycles is 55

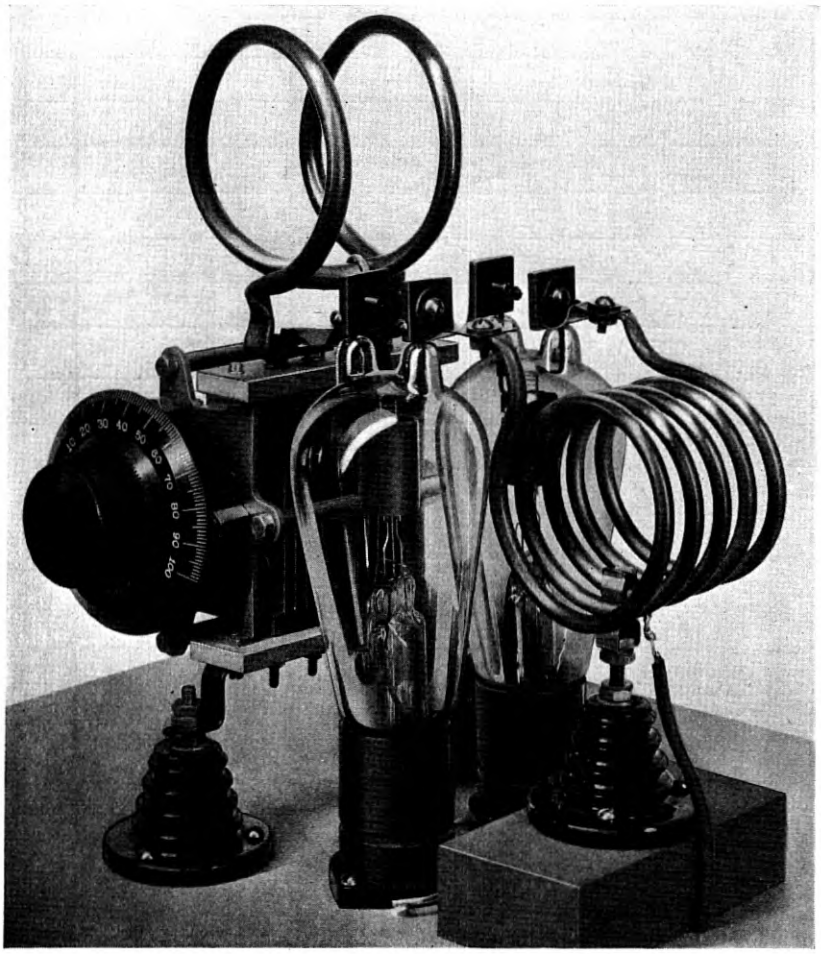

Fig. 9—A typical push-pull circuit for use at ultra-high frequencies. A circuit of this type was used to obtain the data shown in Figs. 8 and 11.

watts per tube and the efficiency 50 per cent, corresponding roughly to the 86 watts output at 47 per cent efficiency for the tube of Fig. 5, the output at 200 megacycles is 34 watts at 33 per cent efficiency as compared to less than 10 watts at an efficiency of only a few per cent

for the other tube. At 300 megacycles the higher frequency tube gives an output of 13 watts, while the tube of Fig. 5 no longer oscillates. This difference in behavior is due primarily to the decreases in transit time and in circuit losses, and to the more nearly optimum ratio existing between the inter-electrode capacitances. It is due to a considerably less extent to the increase in the frequency limit set by circuit resonance.

### *A Still Further Departure in Construction*

The tube illustrated in Fig. 10 represents a still further departure from conventional construction with a corresponding increase in the

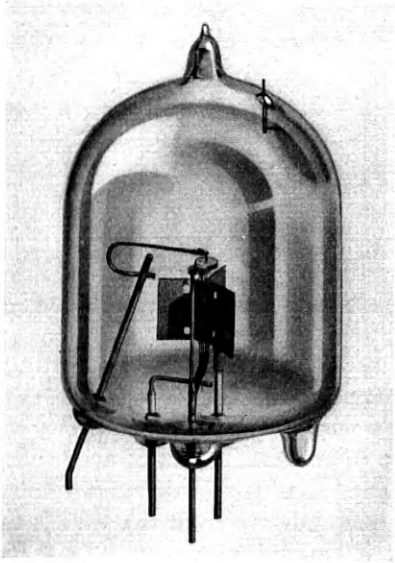

Fig. 10—This tube will oscillate at frequencies up to 740 megacycles. Note the absence of the usual press and the extremely small size of the elements.

frequency limit. Fay and Samuel report an output of 6 watts per tube at 500 megacycles and a frequency limit of 740 megacycles. Unusual features of the design are: The complete elimination of the usual press, the close spacing of the leads and the special construction of the tube elements, particularly the grid, made necessary by their small size. The grid is in the form of a number of straight wires (parallel and equidistant from the axial filament) supported by cooling collars at each end. The plate, in spite of its small size, can dissipate 40 watts with safety.

The inter-electrode capacitances of this tube are:

| | |
|---|---|
| Plate to grid | 1.8 $\mu\mu$f |
| Grid to filament | 1.0 $\mu\mu$f |
| Plate to filament | 0.75 $\mu\mu$f |

The output and efficiency curves, shown in Fig. 11, were obtained with two tubes in a push-pull circuit. The efficiency of the tube is

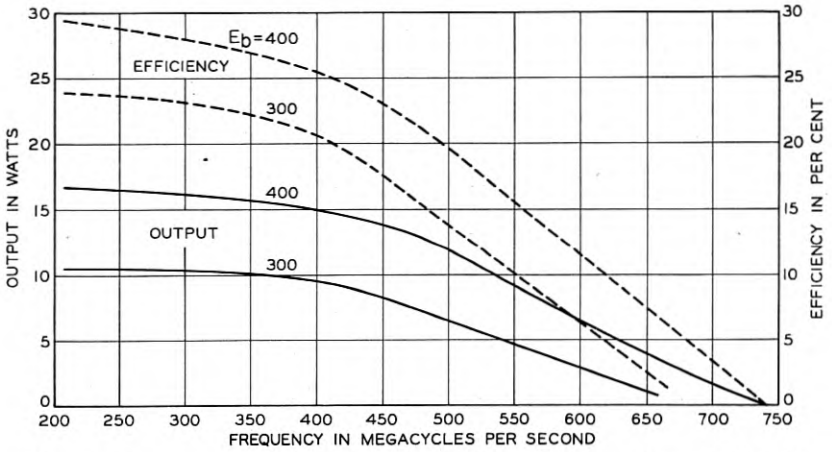


Fig. 11—Output and efficiency curves for 2 tubes of the type shown in Fig. 10.

about 28 per cent at 300 megacycles, contrasting with only 18 per cent for the previously discussed tube, while the output is only 8 watts as compared to 13 watts. The higher efficiency suggests that above 300 megacycles two of the smaller tubes are preferable to one of the large tubes. The fact that the limiting frequency varies with the applied anode potential indicates that the transit time effect is largely responsible for the decreased efficiency. This suggests modulation difficulties if the tube is used near the upper limit of the frequency range. The outputs and efficiencies in the 400- to 600-megacycle range, although low, represent a substantial increase over the usual values obtained at these frequencies.

The relatively low efficiency at 200 megacycles is due to insufficient filament emission. In order to maintain space charge conditions near the filament, at high frequencies, the electron emission must be large enough to supply not only the actual electron current to the plate but also the charging currents to the grid-filament and plate-filament capacitances. In a tube of this small size an increase in the filament emission is possible only by an impractical and unwarranted increase in the filament heating current.

The logical extension of these principles to increasingly high frequencies requires the use of closer and closer inter-electrode spacings. Severe mechanical difficulties are encountered. Curiously enough the limiting factor in the power dissipating ability of the tube turns out to be the grid temperature rather than the temperature of the plate as might be expected. This comes about because of the required close grid-filament spacing, and makes necessary the adoption of some method of cooling the grid. One of the writers has constructed a series of tubes in which the grid is a tungsten helix, each turn of which is attached to a common cooling fin projecting through a slot in the plate. This construction simplifies the mechanical problems involved and provides ample grid cooling. Two of these tubes are shown in Fig. 12.

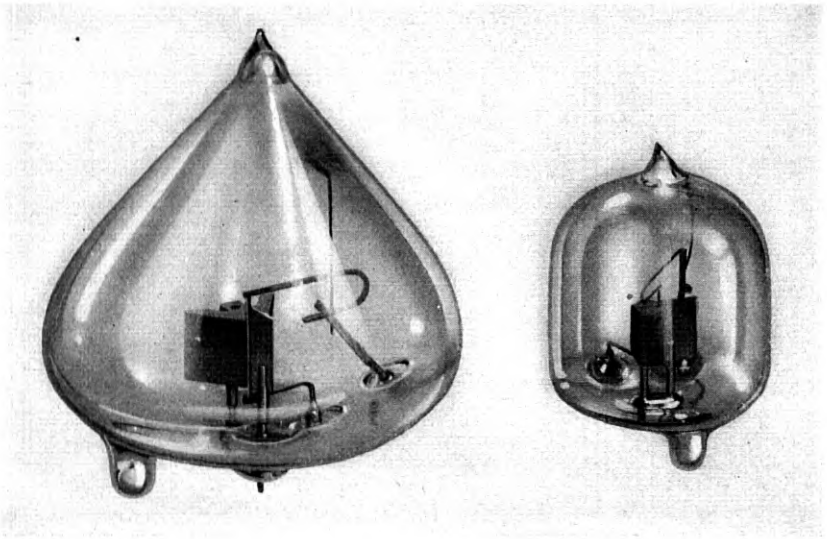

Fig. 12—These tubes represent a further extension in design according to the principles under discussion. The smaller one will oscillate at 1200 megacycles per second.

The larger one will deliver 10 watts at 670 megacycles with an efficiency of 20 per cent and the smaller one will deliver one watt at 1200 megacycles with an efficiency of 10 per cent. These tubes are in no sense commercial, the results representing the limit that has been obtaned by specially constructed tubes under controlled laboratory conditions at voltages and currents above those for which the tube would have a long life. With further advances it is reasonable to expect that outputs of this sort will be commercially realizable and that the frequency

range of the negative grid oscillator can be extended beyond 1200 megacycles.

*Tubes for Receiving Purposes*

For receiving purposes where large outputs are not needed, the ultra-high-frequency requirements may be met by shrinking all the tube dimensions in proportion to the desired wave-length. Tubes constructed by Thompson and Rose [3] based upon this principle are shown in Fig. 13 compared in size with the conventional receiving type tube.

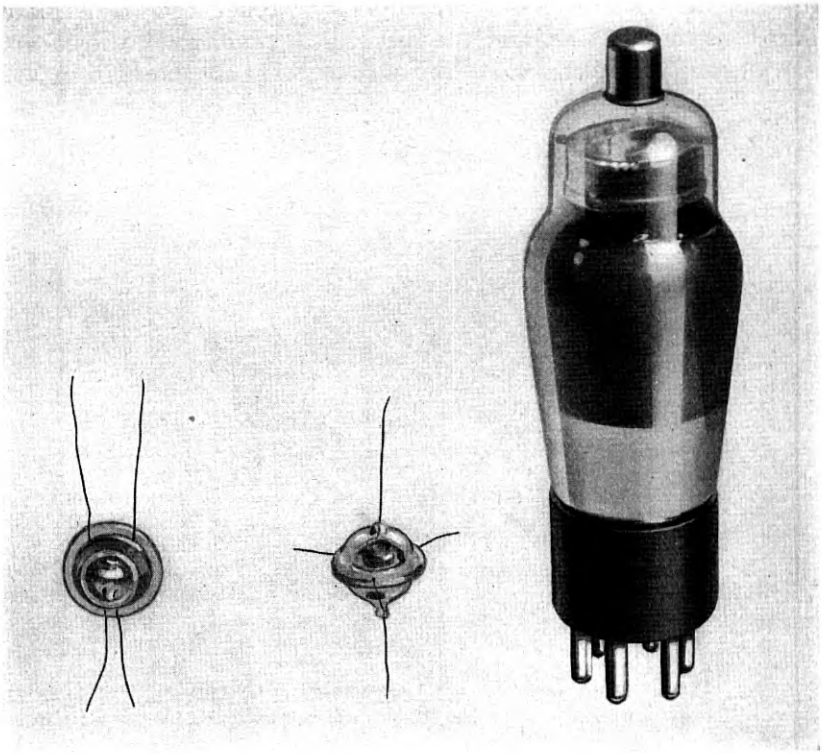

Fig. 13—Receiving tubes of extremely small dimensions. A conventional receiving type tube is shown at the right for comparison.

These tubes make use of parallel plane electrodes, the cathode being oxide coated and indirectly heated, and the grid being in the form of a mesh. A cross-section view of the triode is shown in Fig. 14. This tube will oscillate at frequencies up to 1000 megacycles in miniature replicas of the customary circuits used at longer wave-lengths. A photograph of a complete oscillator is shown in Fig. 15 and the circuit diagram in Fig. 16.

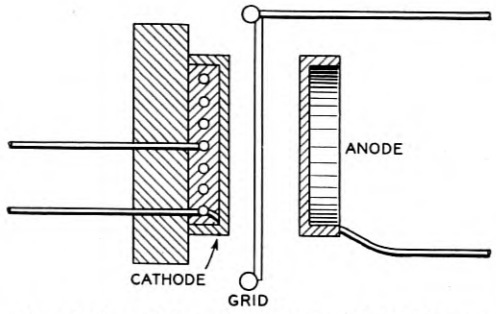


Fig. 14—A sectional view of the triode developed by Thompson and Rose.

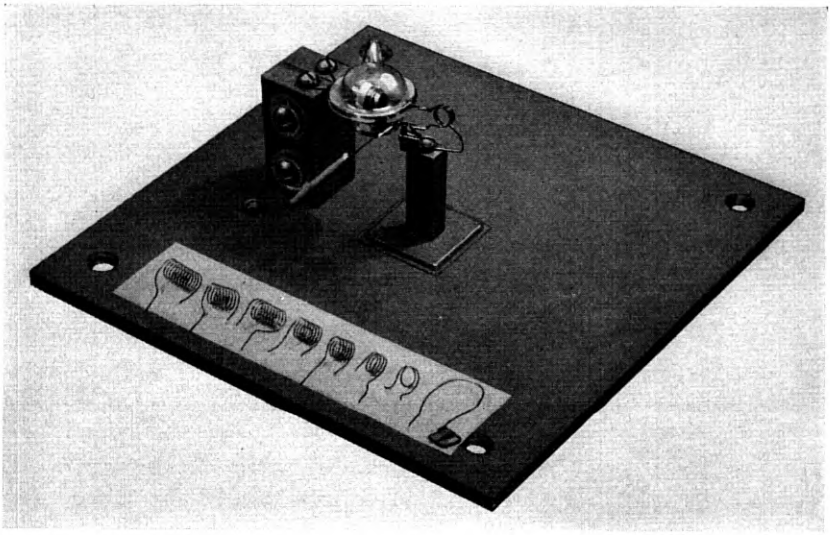

Fig. 15—A complete oscillator using a miniature receiving tube.

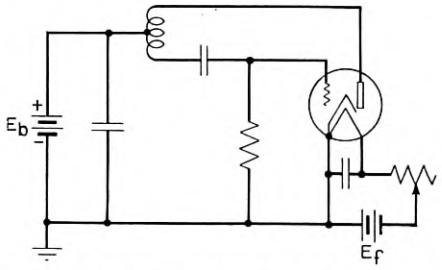


Fig. 16—The circuit diagram of the oscillator shown in Fig. 15.

*Factors Limiting Ultra-High Frequencies*

The limiting factor in the continued extension of the negative grid oscillator to higher and higher frequencies appears to be the dependence of physical size and output power on the operating frequency range. This dependence is well illustrated by the comparison in Fig. 17 of

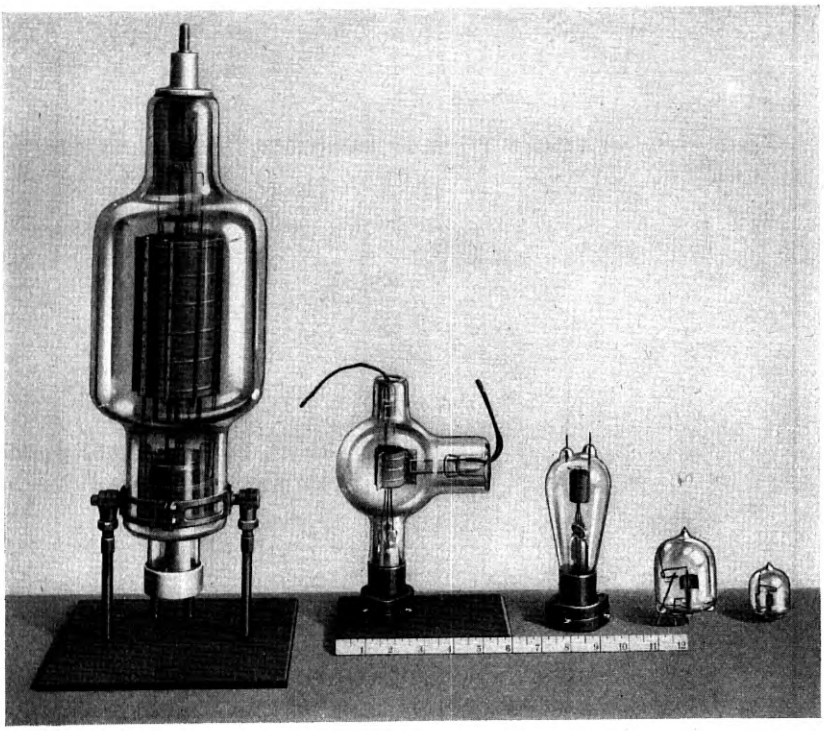

Fig. 17—Some of the tubes, previously discussed, compared in size with a lower frequency tube on the left.

some of the tubes so far discussed with a standard one-kilowatt tube for which the frequency limit is approximately 75 megacycles. Dimensional considerations indicate that the linear dimensions of a series of tubes of optimum design must be decreased in proportion to the operating wave-length. Since the heat dissipating ability depends upon the surface area of the plate, the output (assuming the same efficiency) will decrease as the square of the wave-length. That this is approximately true for the tubes shown in Figs. 5, 7, and 10 may be seen by reference to Fig. 18 where the outputs as a function of frequency are plotted on a logarithmic scale. The sloping lines are for

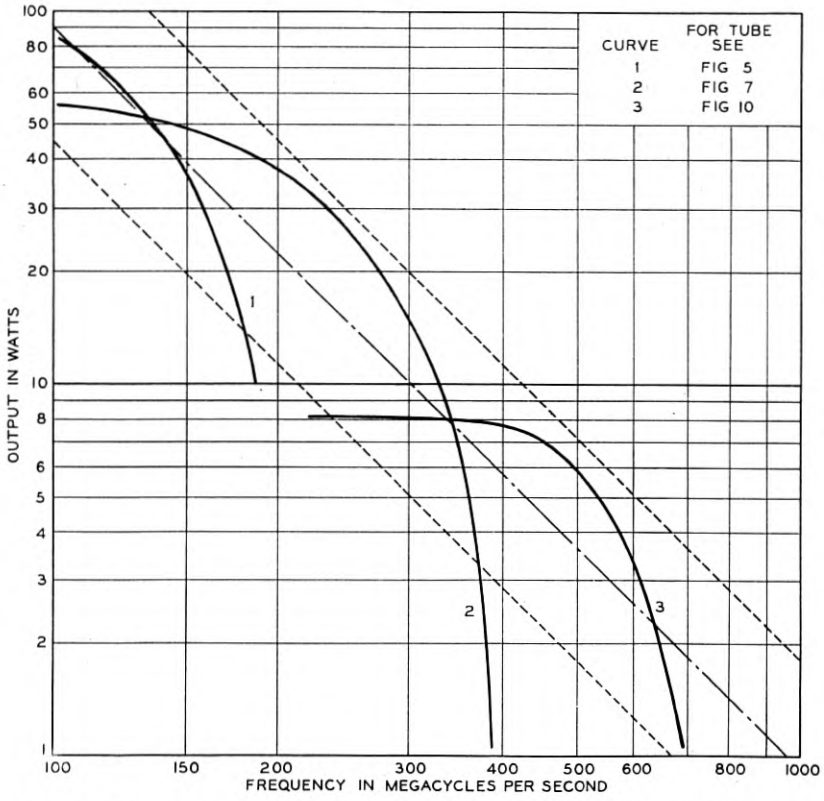


Fig. 18—A comparison plot of the outputs obtained with the tubes of Figs. 5, 7, and 10. The sloping lines are fixed values of the ratio of output to the square of wave-length.

different values of the ratio of output to the square of the wave-length. With radiation-cooled tubes patterned after those illustrated, outputs of only a few tenths of a watt at 3000 megacycles can be expected. If larger outputs are to be obtained, innovations in tube design must be made.

### Positive Grid (Barkhausen) Oscillator

As first reported by Barkhausen and Kurz [7] in 1920, oscillations at frequencies greater than 300 megacycles can be produced by most high-vacuum triodes having symmetrical cylindrical structures when the grid is operated at a fairly high positive potential and the plate is held at or near the cathode potential. When so used they are variously known as oscillators of the Barkhausen and Gill-Morell [8] types after the earliest experimenters, or oscillators of the positive grid or re-

tarding field type to designate the arrangement of the electrode potentials. The relative ease with which such oscillations can be obtained at frequencies above 300 megacycles by the use of conventional tubes, and the widespread interest in this frequency range for communication purposes, have led to the appearance of a large number of papers on the experimental and theoretical aspects of such operation.

With the positive grid oscillator there are found to exist preferred frequencies of operation fixed by the electrode spacings and the applied electrode potentials. For the lowest preferred frequency mode of oscillation the relationship is such that the period of one complete oscillation is approximately equal to the total transit time of an electron which fails to strike the grid on its first transit, is retarded and finally turned back by the plate potential, and again missing the grid, returns to the cathode. Under these conditions the relationship,

$$\frac{E_g}{n^2} = \text{constant}$$

is found to hold approximately, where $E_g$ is the applied grid potential, $n$ is the frequency, and the constant is a function of the tube geometry. Other high-frequency modes of oscillation can be obtained. One of these is particularly easy to excite if the grid of the tube is in the form of a simple helix. The important rôle played by the electron transit time in determining the frequency of the positive grid oscillator contrasts sharply with the minor rôle it plays in determining the frequency of the negative grid oscillator.

For maximum output it is necessary to adjust the tuning of the external circuit to correspond to the preferred frequency fixed by the applied electrode potentials. The relative dependence of the frequency upon the circuit tuning and on the applied electrode potentials varies greatly with the design of the tube. In any case the improper adjustment of either parameter results in a marked decrease in output. In general it appears that the better the tube design and the higher the operating efficiency the greater will be the dependence of frequency upon circuit tuning and the less will be its dependence upon the applied electrode potentials.

The most efficient operation of the positive grid oscillator is obtained when the space current is limited by the cathode emission, as contrasted with the most efficient operation of the negative grid oscillator when the current is limited by space charge. Not only must the space current be emission limited but it must have a fairly critical value. This makes it necessary to adjust the cathode temperature critically. Since the cathode emission characteristics are apt

to change with time, frequent readjustments of the cathode temperature are usually required.

No completely satisfactory and generally accepted theory of the positive grid oscillator has as yet been given. Many theoretical papers dealing with the mechanism of oscillation have been published. Some of these papers resort to pictorial explanations which, from their very nature, must leave out certain basic factors. Readers interested in a résumé of the various theories are referred to the excellent review by Megaw [13] and to the original papers. It is now recognized that any accurate theory must be based upon a general consideration of all the forces acting upon the electrons in their flight between the electrodes. This may take the form of either a particular solution of the classical electromagnetic equations for the conditions within the tube or an analysis of the energy contributions due to individual electrons in their passage across the inter-electrode space.

### *Construction of a Positive Grid Tube*

A representative positive grid tube of current design described by Fay and Samuel [2] before the International Scientific Radio Union is shown in Fig. 19. This tube differs from the conventional negative grid tube primarily in the construction of the grid and in the arrangement of the leads. While designed primarily for use in the frequency range from 500 to 550 megacycles, it illustrates the general problems encountered in the construction of the positive grid oscillator of this type for any frequency range.

The grid consists of a number of parallel wires supported by cooling collars at each end, the so-called squirrel cage construction. It will withstand 150 watts heat dissipation safely, and provides a minimum of circuit inductance and resistance. The grid diameter is fixed by the frequency for which the tube is designed and by the desired operating potential, such that the relationship

$$d_g = \frac{K_1\sqrt{E_g}}{n} \tag{1}$$

is approximately satisfied, where $d_g$ is the diameter of the grid, $K_1$ is a constant, $n$ the frequency, and $E_g$ the applied grid potential.

An indefinite increase in output at a fixed frequency by the simultaneous increase in the grid diameter and in the applied grid potential is not possible because of the limited permissible grid dissipation per unit area. The optimum grid current is found to follow roughly a 3/2

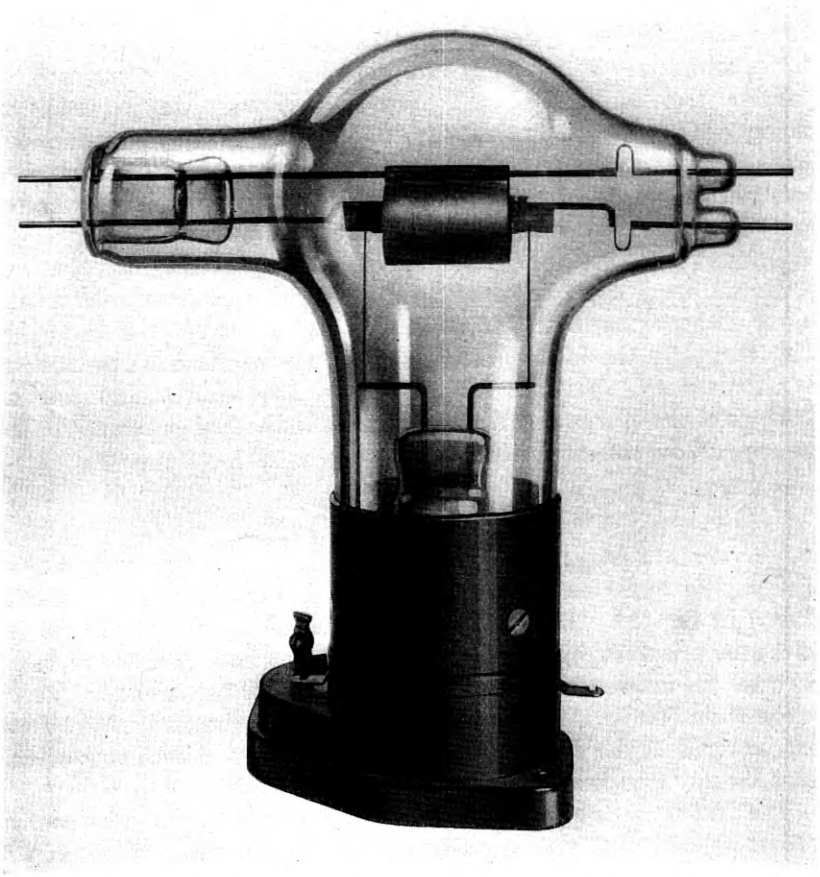

Fig. 19—A positive grid oscillator designed for the frequency range from 500 to 550 megacycles.

power law, that is

$$I_g = K_2 \frac{E_g^{3/2}}{d_g}, \tag{2}$$

so that the grid power will increase as the fourth power of the grid diameter while the grid area and hence the heat dissipating ability only increases as the first power of the diameter. Because of this an upper limit in output exists, fixed by the maximum permissible heat dissipation per unit area for the grid structure. The optimum grid diameter will vary directly with the wave-length for which the tube is designed, and if the ratio of the grid length to its diameter is main-

tained constant, the maximum available power output (assuming the same efficiency) will vary as the square of the desired wave-length.

*Circuit of a Positive Grid Oscillator*

In Fig. 20 is shown a diagram of a positive grid tube of the straight-wire-grid type and its associated circuit. Tuned circuits, in this case in the form of so-called Lecher systems, are connected between the grid and plate leads, extending approximately a half wave-length (30 cm) beyond the lead seals. Because of the existence of preferred

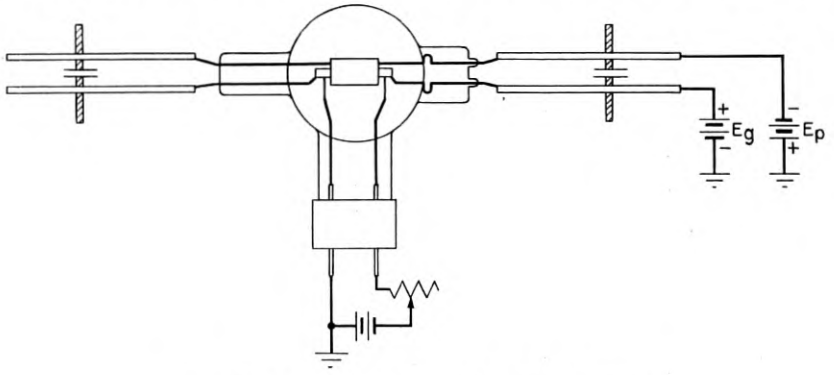


Fig. 20—A typical positive grid oscillator circuit.

frequencies of operation fixed by the potentials applied to the tube electrodes, distributed-constant circuits, if used, may be operated at frequencies corresponding to harmonic modes of oscillation. In this case the length of the leads within the tube envelope has been adjusted so that the glass seals come at or near potential nodal points for the Lecher systems of which the leads form a part. This minimizes dielectric losses in the glass. The effective paralleling of the two sets of leads greatly reduces the resistance losses, while the balanced arrangement decreases radiation losses. Strict attention to these details is required because of the already low efficiency of the mechanism of generation.

*Characteristics of Positive Grid Oscillators*

The dependence of output and anode efficiency on frequency is shown in Fig. 21. These data were taken by adjusting the circuit tuning, filament current, and the grid and plate potentials to their optimum values for each frequency. The curve showing the grid voltage will be observed to follow equation (1) above, at least roughly, and a similar correspondence will be observed between the curve for

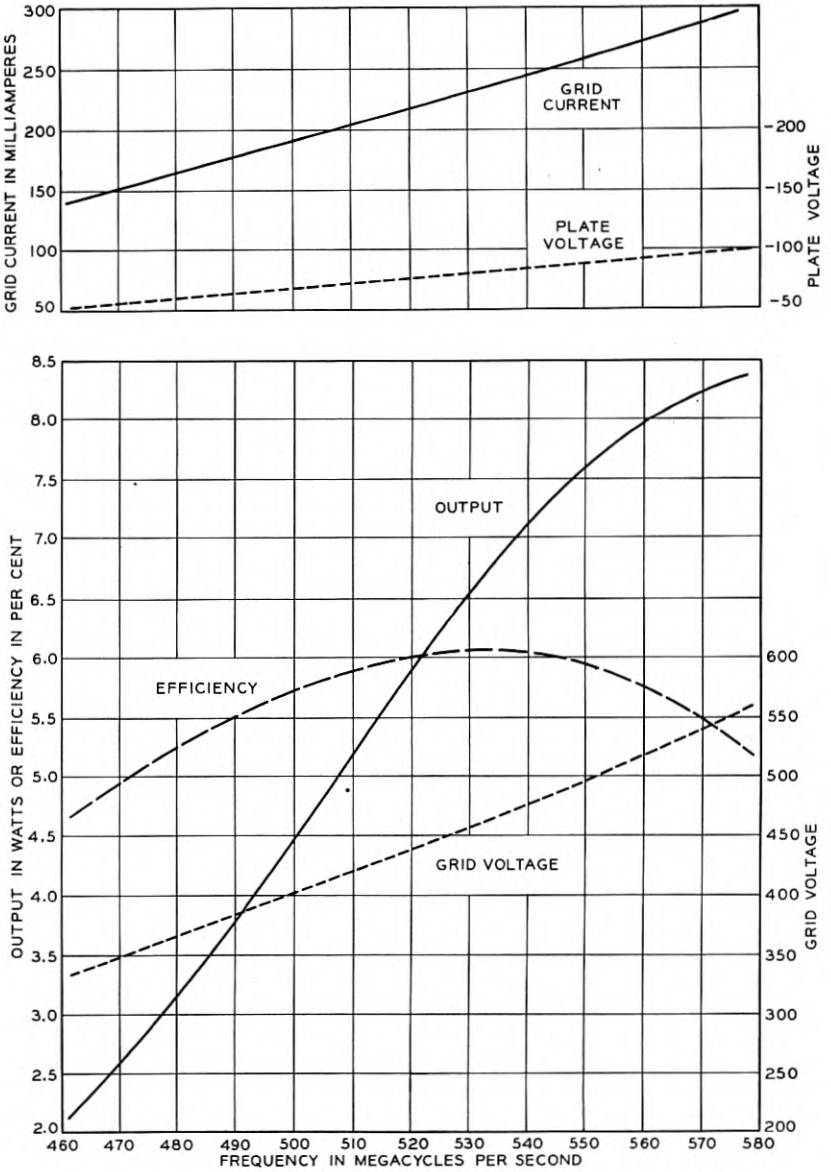


Fig. 21—Output and efficiency curves for the tube shown in Fig. 19.

the grid current and equation (2). Some variation in the required negative plate potential is observed. A maximum efficiency will be noted at a frequency of approximately 530 megacycles. The output, however, continues to increase with increasing frequency, the limit in output as well as in frequency being set by the safe grid dissipation. The outputs over the 500- to 600-megacycle range vary from 4.5 to 8 watts, comparing with outputs from 6 to 3 watts for the negative grid tube. The low efficiencies of 5 to 6 per cent are to be compared with the somewhat higher efficiencies of 19 to 11 per cent for the negative grid type tube.

The influence of the grid voltage on the frequency and on the output and efficiency is shown by the curves in Fig. 22. These data are for a

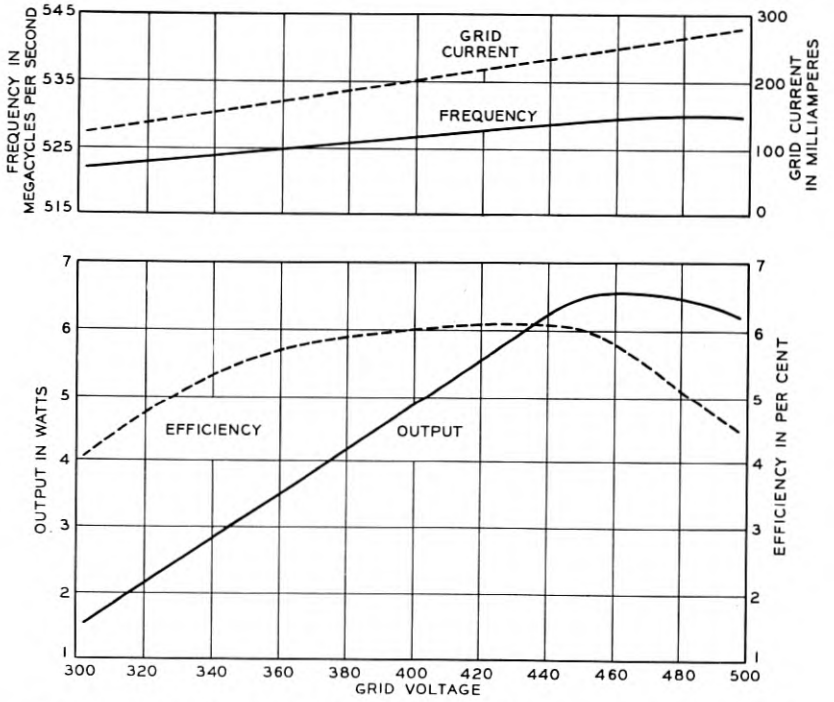


Fig. 22—Result of variation of grid voltage, tuning fixed, on tube of Fig. 19.

fixed circuit tuning, the grid voltage being adjusted to the values indicated. This corresponds to the condition that might obtain if a grid voltage modulation scheme were to be used. The lack of linearity of the output curve and the large shift in frequency indicate that amplitude modulation by this method would be unsatisfactory. The

frequency shift, large as it is, is much less than the shift observed in tubes of poorer design and correspondingly lower efficiency.

The dependence of the output and efficiency as well as frequency upon the grid current is shown in Fig. 23. These data were taken with

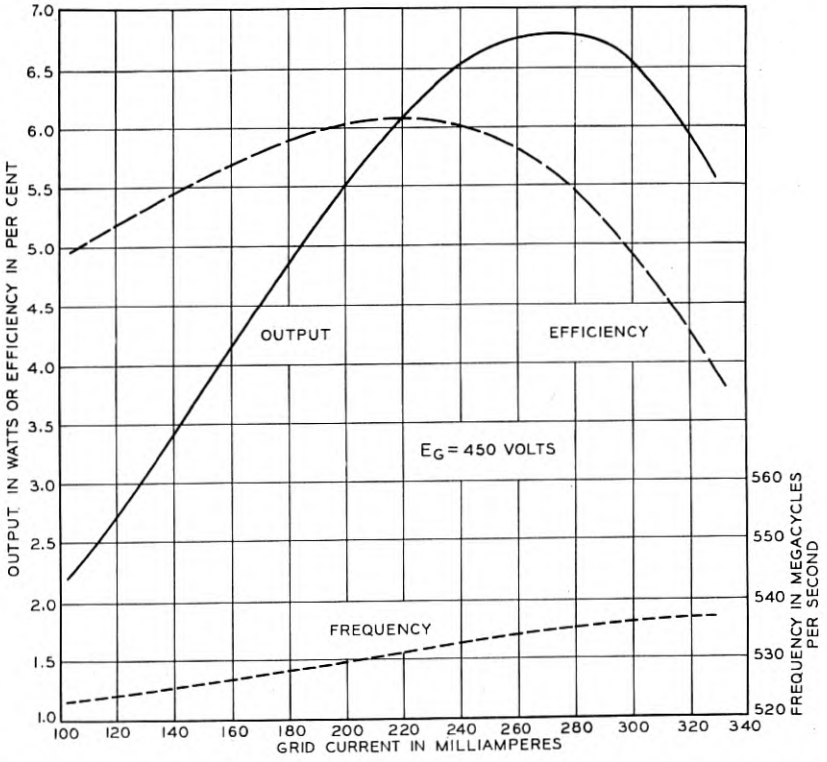


Fig. 23—Result of variation of grid current, other adjustments fixed, on tube of Fig. 19. $f$ = 532 megacycles.

a grid potential of 450 volts and a fixed circuit adjustment corresponding roughly to a frequency of 532 megacycles. The current to the grid was varied by adjusting the temperature of the filament. The maxima observed in both the output and the efficiency correspond to conditions for which the grid current is limited primarily by the available emission rather than by space charge. As conditions corresponding to complete space charge are approached the output and efficiency fall off rapidly. The limit on the permissible grid dissipation prevents the extension of these curves to the condition of complete space charge. Because of this dependence of output on grid current, the adjustment of filament temperature is extremely critical.

At lower frequencies the efficiency of operation of a correctly designed positive grid tube is substantially the same as that exhibited by this tube. The negative grid oscillator on the other hand, as has been shown, increases both its output and efficiency rapidly with decreasing frequency. The positive grid oscillator is, therefore, at an increasing disadvantage at lower frequencies. With the present state of development, the negative grid oscillator will give larger outputs with higher anode efficiencies at all frequencies less than about 300 megacycles.

For frequencies much higher than 600 megacycles, it is found that the power input requirements for efficient operation of tubes having grids of the straight wire type are in excess of that which can be tolerated in the grid structures. Operation at very much less than optimum input results in considerable decrease in output as indicated in Fig. 25.

*Spiral Grid Barkhausen Tubes*

If the grid of a Barkhausen oscillator is in the form of a simple helix, oscillations at frequencies greater than those predicted by the relationship of equation (1) are readily obtained. When so constructed they are called spiral grid Barkhausen tubes. Some experimental models are shown in Fig. 24. The tubes used in the Lympne to St.

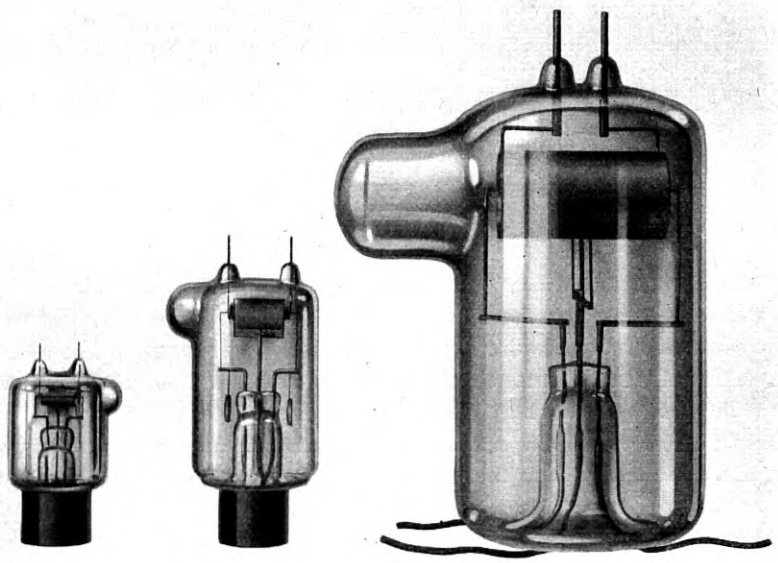

Fig. 24—Three optimum positive grid oscillators of the spiral grid type. Smallest tube designed for 2500 megacycles, largest for 500 megacycles.

Inglevert "micro-ray link" are of this general type.[14] Such tubes have been used to produce oscillations up to 3000 megacycles.

Because of the fact that the severe limitation on the optimum grid diameter is modified by the presence of a tuned circuit within the tube formed by the helical grid helix, tubes of this type are particularly useful at frequencies above 600 megacycles. The former restriction on grid diameter is replaced by the requirement that the expanded length of the grid spiral be approximately 1.24 times the wave-length at which the maximum output is required, and that there be a correct proportioning of the other dimensions of the tube. The dependence of wave-length on the grid wire length is illustrated by the experimental data in Table I covering a frequency range from 460 megacycles

TABLE I

DEPENDENCE OF WAVE-LENGTH ON GRID WIRE LENGTH FOR BARKHAUSEN OSCILLATOR WITH HELICAL GRID

| Grid Wire Length in Cm. | Optimum Wave-length in Cm. | Ratio |
|---|---|---|
| 16.3 | 13.5 | 1.21 |
| 18.6 | 18.6 | 1.00 |
| 19.7 | 14.5 | 1.36 |
| 20.4 | 18.0 | 1.13 |
| 21.4 | 17.5 | 1.22 |
| 21.3 | 25.0 | 0.85 |
| 22.3 | 20.2 | 1.10 |
| 25.2 | 18.6 | 1.35 |
| 30.6 | 25.0 | 1.22 |
| 32.0 | 23.6 | 1.36 |
| 32.0 | 25.5 | 1.25 |
| 33.4 | 25.0 | 1.33 |
| 42.6 | 29.0 | 1.47 |
| 42.6 | 29.5 | 1.44 |
| 42.6 | 30.2 | 1.41 |
| 42.6 | 30.7 | 1.38 |
| 53.2 | 43.5 | 1.22 |
| 80.0 | 65.0 | 1.23 |
| Average | | 1.24 |

to 2220 megacycles. Graphic evidence of the independence of shape is shown by the largest and the smallest tube shown in Fig. 24 for which the dimensional ratios are nearly the same. The largest tube delivers several watts at 500 megacycles, while the smallest one delivers only a few tenths of a watt at 2500 megacycles. The efficiency in both cases is about one per cent. These dimensional considerations lead to the conclusion that there exists a maximum output at any given wave-length for a tube of a given design and that this output is proportional to the square of the optimum wave-length. From this it appears that the only advantage offered by the spiral grid tube over the other type is the simplification in mechanical design which

permits the construction of rigid grid structures capable of high energy dissipation for the higher frequency range.

The external tuned circuit for the higher frequency mode of oscillation takes the form of a Lecher system connected between the two grid terminals. When so connected the dependence of frequency upon circuit tuning is pronounced, as contrasted with the negligible dependence observed if the Lecher system is connected between the plate and the grid. When oscillating in the higher-frequency mode the spiral grid tube shows only a comparatively small dependence of frequency on grid potential and this may be compensated by a proportional change in plate potential. This, coupled with the fact that the output increases rapidly with increasing grid potential, makes it possible to apply various schemes of amplitude modulation. Characteristics of the type shown in Figs. 21, 22, and 23 for the straight-wire-grid tube cannot be taken except for a limited portion of the range due to the inability of the grid to dissipate the energy required in the upper portion of the grid voltage or grid current ranges.

While the spiral grid tube will also oscillate in the lower-frequency mode, its efficiency and output are considerably lower than the corresponding values for the straight-wire-grid tube previously discussed. Its field of usefulness is, therefore, largely limited to the higher frequency mode of oscillation in the frequency range above 600 megacycles.

## The "Magnetron" Oscillator

The "magnetron" in its simplest form consists of a cylindrical diode or 2-electrode tube, with a uniform magnetic field in the direction of the electrode axis. The original type of tube has been largely superseded for ultra-high-frequency generation by the so-called split-plate magnetron, first used by Okabe,[18] in which the cylindrical anode is divided longitudinally into two (or more) segments to the terminals of which is connected the tuned circuit. Such a tube is shown in Fig. 25.

In the frequency range from 300 to 600 megacycles the split-plate magnetron compares favorably with the negative grid tube both in output and in anode efficiency. Its use has been limited because of the complicating factor of the magnetic field, and the attending modulation difficulties. For frequencies higher than 600 megacycles the magnetron provides larger outputs than those so far reported by other means. It has been used at frequencies up to 30,000 megacycles, a value well above that so far reported for any other type of vacuum tube.

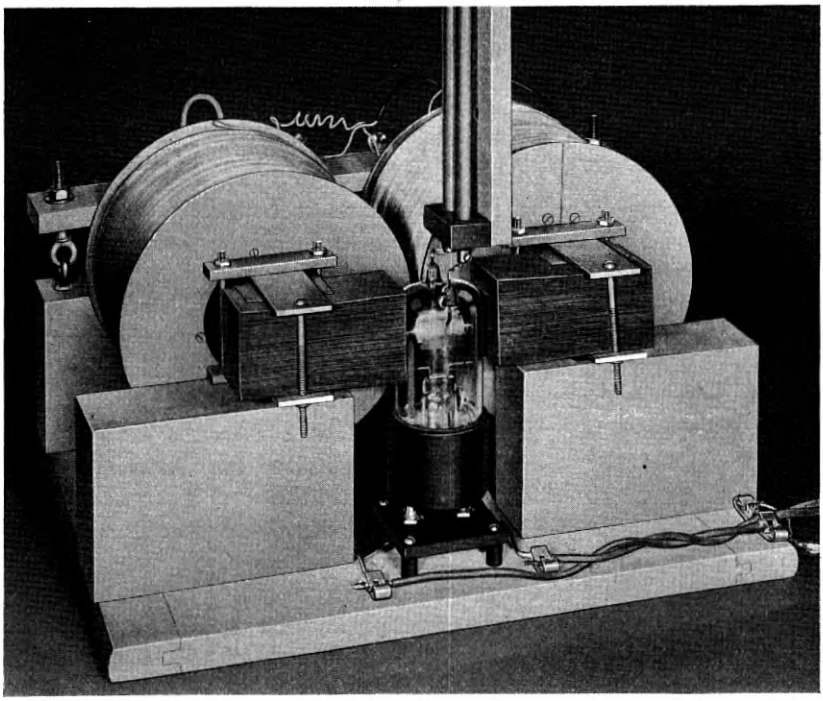

Fig. 25—An experimental model of the split-plate magnetron showing a possible arrangement of the magnetic field.

The magnetron depends for its operation upon the curvature of the electron orbits produced by the magnetic field. As first shown by Hull [15] in 1921, a critical field exists beyond which the anode current falls off rapidly to zero. This field is given by the relationship

$$H = \frac{6.72}{R}\sqrt{V}, \tag{3}$$

where $R$ is the anode radius and $V$ is the potential of the cylindrical anode with respect to an axial filament. Although the original magnetron of Hull and Elder made use of variations in the magnetic field in its operation as a generator, it was soon discovered that oscillators could also be produced with steady fields by two somewhat different mechanisms. The one, first pointed out by Habann,[16] makes use of a negative resistance effect observable in the static characteristics and the other, first described by Žáček,[17] involves the electron transit time in a way quite analogous to the way in which it is involved in the positive grid triode. Both mechanisms have been used to produce oscillations at ultra-high frequencies.

*Negative Resistance Type*

Data reported by McArthur and Spitzer [1] on a split-plate magnetron tube are illustrative of the negative resistance type of behavior. Static characteristics taken by varying the potential of one anode with the magnetic field held constant for different values of the potential on the other anode are shown in Fig. 26. The pronounced negative

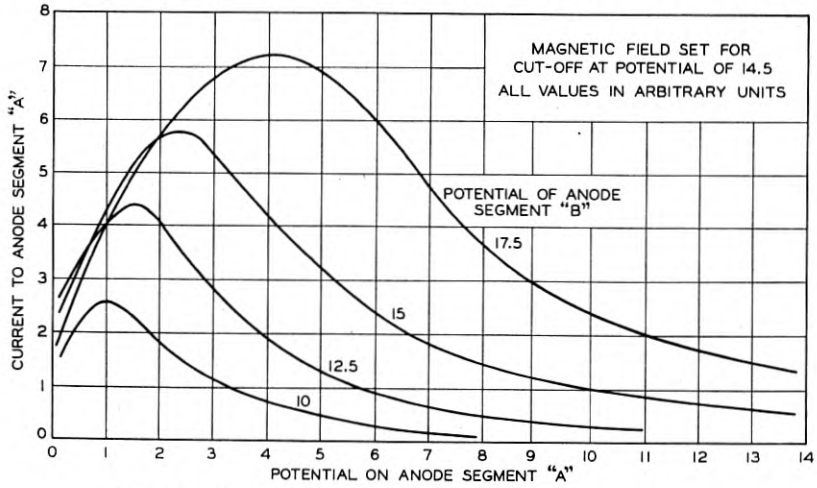


Fig. 26—Static characteristics of a split-plate magnetron.

resistance effect is obvious. This negative resistance characteristic can be utilized in producing oscillations.

Output and efficiency curves for this tube as an oscillator are shown in Fig. 27. These data were obtained by connecting a "tank"

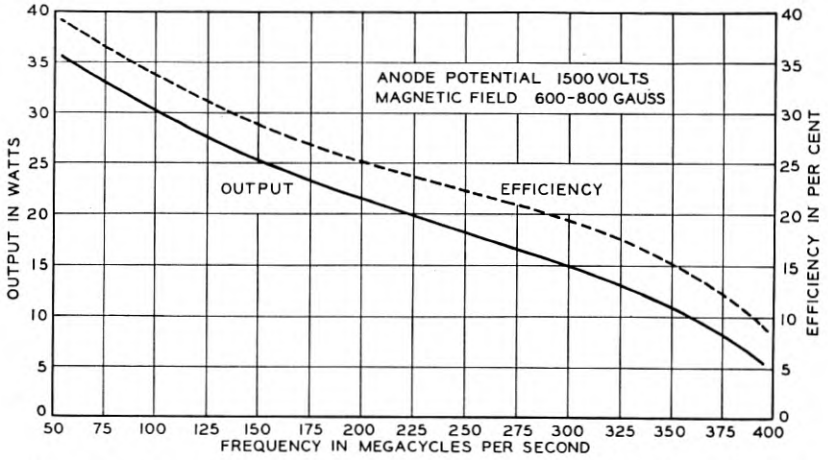


Fig. 27—Output and efficiency curves at different frequencies for split-plate magnetron.

circuit, tuned to the desired frequency, across the two anodes, as shown in Fig. 28. Each anode delivers energy to the oscillating circuit during alternate half-cycles, so that in effect, it is equivalent to a push-pull oscillator. The limiting frequency as set by the inter-electrode capacitances and lead inductances (corresponding to the similar limit for the negative grid tube) is 450 megacycles. The de-

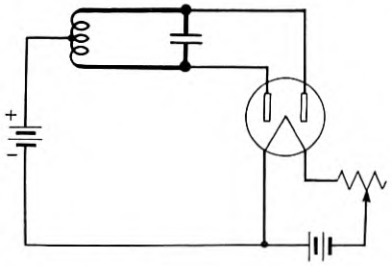

Fig. 28—Oscillation circuit of split-plate magnetron. Data for Fig. 27 taken in this type of circuit.

crease in output before this limit is reached is due to resistance and radiation losses and to the effect of electron transit time.

The magnetron, as contrasted with the negative grid tube, will oscillate with circuits having a high decrement. However, for its most efficient operation the effective anti-resonant impedance of the tuned circuit when loaded must be approximately 10 times the value required by a triode with the same anode dimensions. The load resistance that can be obtained at high frequencies is only a fraction of this value, so that the efficiency becomes increasingly less with higher frequencies. A further limitation is due to the fact that the electron current is concentrated on only a small part of the anode surface. This reduces the safe anode dissipation unless the anode is designed to have a high thermal conductivity. Because of these limitations, the ratio of output to the inter-electrode capacitance may be only slightly more favorable than the corresponding ratio for a triode of the same anode dimensions.

*Type Depending on Electron Transit Time*

When the magnetic field of a split-plate magnetron is adjusted to near the critical value given by equation (3), oscillations can be produced whose frequency will depend primarily upon the time of flight of electrons between filament and anode in a way closely resembling the behavior of the positive grid oscillator in its lower frequency mode of oscillation. For best output, the field must be above the critical

value. To fix the time of flight and hence the frequency of oscillation, the magnetic field and plate voltage must be adjusted to certain values roughly expressed by the empirical relationship

$$\lambda H = 13{,}100, \tag{4}$$

where $\lambda$ is the wave-length in centimeters and $H$ is the field strength in gausses, which must also satisfy equation (3). It is found that for best operation the magnetic and electric fields within the tube should not be exactly perpendicular. This lack of perpendicularity may be achieved either by tipping the magnetic field relative to the tube axis or by introducing end plates within the tube and maintaining them at a fixed positive potential.

Kilgore [19] has given complete information concerning this type of oscillator. In Fig. 29, taken from his paper, is shown the dependence

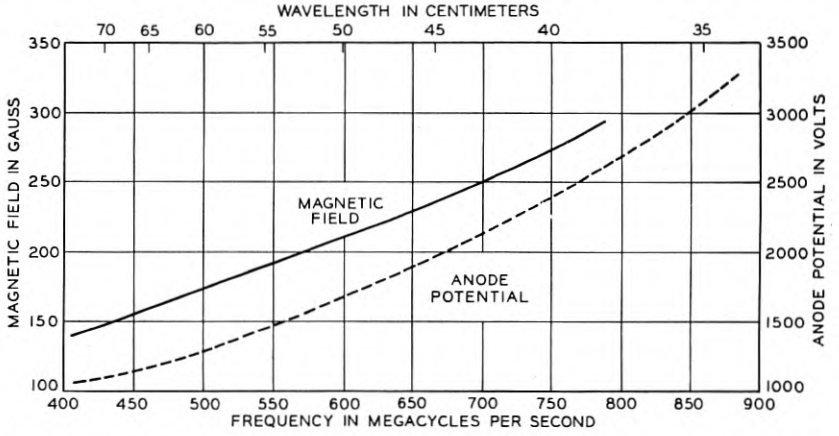


Fig. 29—Relation of magnetic field and anode potential to frequency of oscillation in magnetron oscillator of second type.

of field strength and anode potential on the desired frequency. The existence of a preferred frequency fixed by these values is confirmed by the data in Fig. 30 relating the output and wave-length with the length of the attached Lecher system. The decreased output shown at the second peak is due to the added losses introduced by the extended length of the system. The outputs shown on these curves are not in watts, but represent relative readings of the field strength near the oscillator. With optimum adjustments 7 watts at 715 megacycles is reported, the efficiency being about 8 per cent. The dependence of output and frequency on the applied anode potential

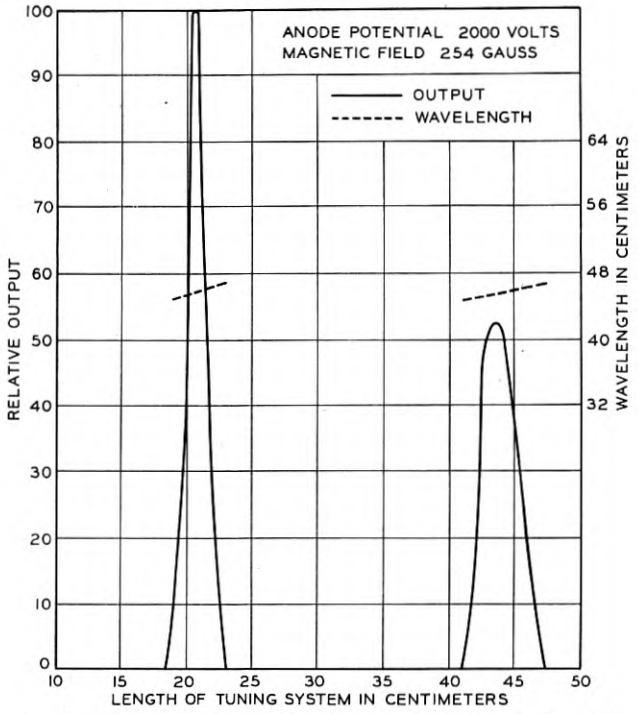


Fig. 30—Relation of wave-length and output to length of tuning system in magnetron oscillator of second type. Anode potential—2000 volts, Magnetic field—254 gausses.

is shown in Fig. 31, and the dependence of frequency on the current in the magnetic field coil in Fig. 32. The importance of the adjustment of the field angle is shown by the data in Fig. 33.

An output of 2.5 watts at 3160 megacycles has been reported by Wolff, Linder, and Braden.[22] They find that the efficiency of the tube is much improved by using end plates in place of the tipped magnetic field. Cleeton and Williams [21] have been able to obtain oscillations at 30,000 megacycles with a magnetron tube.

## Amplification

The use of the conventional thermionic triode as an amplifier greatly exceeds its use as an oscillation generator in communication applications. Its ability to amplify has contributed much more to the development of our present-day long distance communication, whether by wire or by radio, than has its ability to oscillate. The complete utilization of ultra-high frequencies as carrier channels in communica-

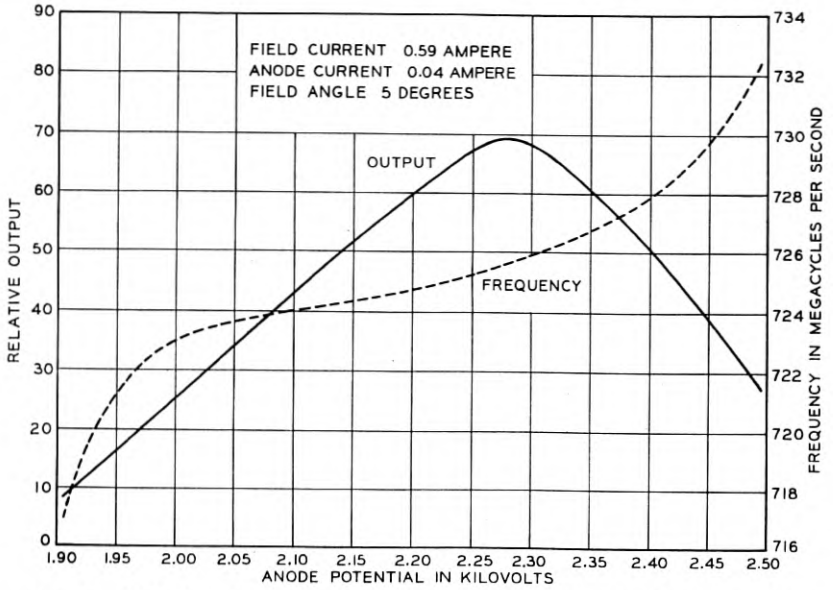


Fig. 31—Relation of output and frequency to anode potential in magnetron oscillator of second type.

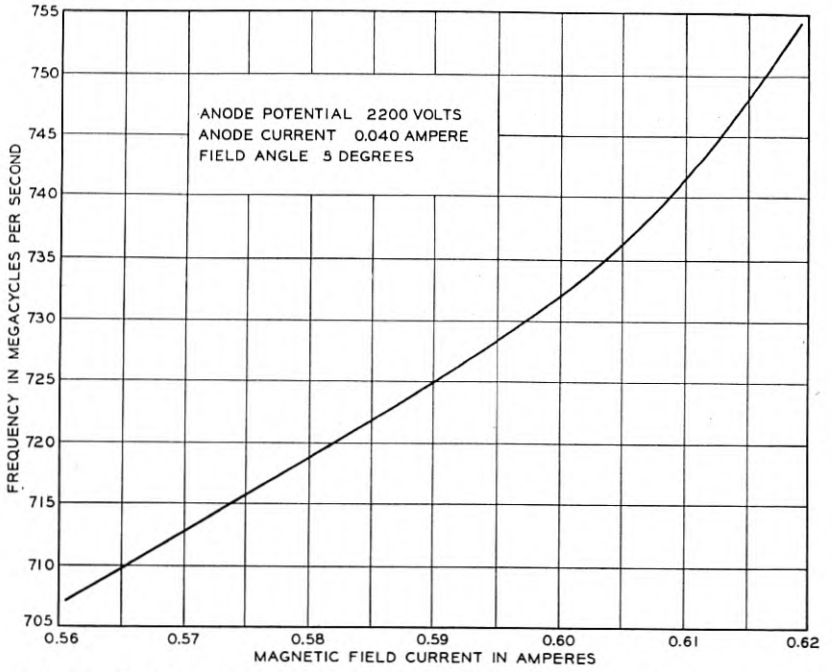


Fig. 32—Relation of frequency to magnetic field current in magnetron oscillator of second type.

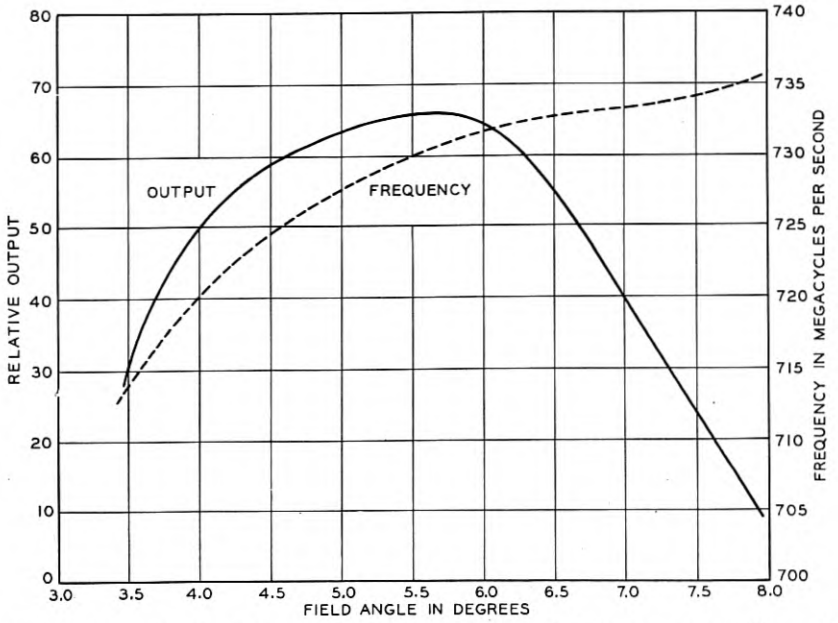


Fig. 33—Relation of output and frequency to field angle in magnetron oscillator of second type.

tion will also, no doubt, be dependent upon the development of suitable amplifiers for this frequency range. Although certain forms of pseudo-amplification are possible with tubes of the Barkhausen and magnetron types, the negative grid triode and multi-element tubes derived therefrom are the only devices available for very high frequencies which will amplify in the sense that the output is an enlarged undistorted replica of the input.

As the frequency of operation of the negative grid triode is increased, difficulties in securing stable operation as an amplifier and in realizing the full gain indicated by the tube constants are encountered. These difficulties, as is well known, are in the main due to the tendency of the amplifier to oscillate or "sing" because of feed-back through the grid-plate capacitance. This may be overcome either by the introduction of a compensating capacitance somewhere in the circuit, so-called neutralization, or by the introduction of an electrostatic shield or screen within the tube envelope between the grid and plate, giving the screen-grid tetrode. Neutralization schemes fail at very high frequencies because of the inductance of the tube leads which makes difficult the correct location of the neutralizing capacity and because of transit-time effects which shift the phase of the needed compensa-

tion. However, conventional screen-grid tetrodes and pentodes are available which function satisfactorily over the major portion of the frequency range covered by the conventional 3-element tube as an oscillator.

For frequencies above approximately 60 megacycles specially designed tubes are required. Because of the similarity in the special frequency requirements, it is expected that there will be found a succession of multi-element tubes for amplification use, each rated for a band of frequencies, patterned after corresponding triode oscillators. The special frequency requirements for the amplifying tube are even more severe than those for the triode oscillator, so that the multi-element amplifying tube will in general cease amplifying at a frequency somewhat lower than the frequency limit of oscillation of the corresponding triode oscillator.

Thompson and Rose [3] have described small screen-grid tubes which will amplify at frequencies of 300 to 400 megacycles. One of these tubes is shown in Fig. 13. Their characteristics are similar to those of the conventional screen-grid tube in many respects. The very great reductions in inter-electrode capacitances, lead inductances, and transit time make possible the construction of receiving circuits using tuned radio frequency amplification at these very high frequencies. The ratio of the frequency limits of the corresponding triode as an oscillator (1000 megacycles) to the frequency at which amplification was reported (400 megacycles) is typical and illustrates the apparently inevitable failure of the amplifier to keep pace with the oscillator in the struggle toward higher and higher frequencies.

## References

*Negative Grid Tubes*

1. E. P. McArthur and E. E. Spitzer, "Vacuum Tubes as High-Frequency Oscillators," *I. R. E. Proc.*, v. 19, 1931, p. 1971. (Bibliography of 28 titles.)
2. C. E. Fay and A. L. Samuel, "Vacuum Tubes for Generating Frequencies Above 100 Megacycles," *I. R. E. Proc.*, v. 22, 1934, p. 679.
3. B. J. Thompson and G. M. Rose, "Vacuum Tubes of Small Dimensions for Use at Extremely High Frequencies," *I. R. E. Proc.*, v. 21, 1933, p. 1707.
4. W. H. Wenstrom, "An Experimental Study of Regenerative Ultra-Short Wave Oscillators," *I. R. E. Proc.*, v. 20, 1932, p. 113.
5. I. E. Mouromtseff and H. V. Noble, "A New Type of Ultra-Short Wave Oscillator," *I. R. E. Proc.*, v. 20, 1932, p. 1328.
6. W. Kroebel, "On the Production of Undamped Oscillations of Wave-Lengths of the Order of a Decimeter in a Reaction Circuit I," *Ann. der Physik*, Series 5, v. 14, 1932, p. 80.

*Positive Grid Tubes*

7. H. Barkhausen and K. Kurz, "The Shortest Waves Obtainable with Vacuum Tubes," *Phys. Zeits.*, v. 21, 1920, p. 1.
8. E. W. B. Gill and J. H. Morrell, "Short Electric Waves Obtained by Valves," *Phil. Mag.*, v. 44, 1922, p. 161.

9. K. Kohl, "Continuous Ultra-Short Electric Waves," *Ergebnisse der exaklen Naturwissenschaften*, v. 9, 1930, p. 275. (Bibliography of 135 titles.)
10. G. Potapenko, "Investigations in the Field of Ultra-Short Electro-Magnetic Waves," *Phys. Rev.*, v. 39, 1932, p. 625; v. 39, 1932, p. 638; v. 40, 1932, p. 988; v. 41, 1932, p. 216. (Many references to the literature.)
11. H. Collenbusch, "Investigations of Very Short Wave Tubes," *Ann. der Physik*, Series 5, v. 13, 1932, p. 191.
12. H. N. Kozanowski, "A New Circuit for the Production of Ultra-Short-Wave Oscillations," *I. R. E. Proc.*, v. 20, 1932, p. 957.
13. E. C. S. Megaw, "Electronic Oscillations," *I. E. E. Jl.*, v. 72, 1933, p. 313. (Bibliography of 47 titles.)
14. H. G. Clavier, "Production and Utilization of Micro-Rays," *Elec. Communication*, v. 12, 1933, p. 3

*Magnetron Tubes*

15. A. W. Hull, "The Effect of a Uniform Magnetic Field on the Motion of Electrons Between Coaxial Cylinders," *Phys. Rev.*, v. 18, 1921, p. 31.
16. E. Habann, "A New Vacuum Tube Generator," *Zeit. f. Hockfreq.*, v. 24, 1924, p. 115.
17. A. Žáček, "A Method for the Production of Very Short Electromagnetic Waves," *Zeit. f. Hockfreq.*, v. 32, 1928, p. 172.
18. K. Okabe, "Production of Intense Extra-Short Electromagnetic Waves by Split-Anode Magnetron," *I. E. E. Jl.*, Japan, 1928, p. 284. See also *I. R. E. Proc.*, v. 17, 1929, p. 652; and *I. R. E. Proc.*, v. 18, 1930, p. 1748.
19. G. R. Kilgore, "Magnetostatic Oscillators for Generation of Ultra-Short Waves," *I. R. E. Proc.*, v. 20, 1932, p. 1741.
20. E. C. S. Megaw, "An Investigation of the Magnetron Short Wave Oscillator," *I. E. E. Jl.*, v. 72, 1933, p. 326. (13 references.)
21. C. E. Cleeton and N. H. Williams, "A Magnetostatic Oscillator for the Generation of 1 to 3-Cm. Waves," *Phys. Rev.*, v. 44, 1933, p. 421.
22. I. Wolff, E. G. Linder, and R. A. Braden, "Transmission and Reception of Centimeter Waves," *I. R. E. Proc.*, v. 22, 1934, p. 555.

*General*

23. F. B. Llewellyn, "Vacuum Tube Electronics at Ultra-High Frequencies, *I. R. E. Proc.*, v. 21, 1933, p. 1531.
24. W. E. Benham, "Theory of the Internal Action of Thermionic Systems at Moderately High Frequencies," Part I—*Phil. Mag.*, March 1928, p. 641. Part II—*Phil. Mag. A*, v. 11, 1931, p. 457.
25. L. Rohde, "On Transmitter Tubes for Producing Waves One Meter Long," *Hochfreq. und Elektroakustik*, v. 40, 1932, p. 3.
26. W. C. White, "The Selection of Types of Vacuum Tubes as Sources of High Frequency Power," *G. E. Rev.*, v. 36, 1933, p. 394.
27. W. H. Wenstrom, "Historical Review of Ultra-Short-Wave Progress," *I. R. E. Proc.*, v. 20, 1932, p. 95. (Bibliography of 54 titles.)

# Horizontal Rhombic Antennas *

By E. BRUCE, A. C. BECK and L. R. LOWRY

The paper discusses the theoretical methods employed by the authors in dimensioning horizontal rhombic receiving antennas. Experimental proof is given of the engineering accuracy of the directivity calculations on which this work is based. There are included brief descriptions of the antenna-to-transmission line coupling circuits and the resistance terminations for rhombic antennas.

## Introduction

AN introductory discussion has been given in a previous paper [1] of a type of antenna which maintains a desirable degree of directivity throughout a broad continuous range of frequencies. This structure was descriptively termed the "diamond-shaped" antenna in that paper, but has since become known as the "rhombic antenna" and will be so designated here.

This paper discusses, in some detail, the theoretical methods employed by the authors in dimensioning horizontal rhombic receiving antennas from data obtained by preliminary surveys of the incident-plane angles of wave-arrival at the proposed receiving site.

Experimental proof is given of the engineering accuracy of the directivity calculations on which this work is based. Checking measurements were made on small-scale rhombic antennas operating at correspondingly short wave-lengths. Confirming data were also obtained on large, adjustable rhombic antennas during the reception of European signals.

The paper also includes a brief discussion of the antenna-to-transmission line coupling circuits employed and, in addition, the resistance terminations, located at the end of the antenna remote from the receiver, used for suppressing standing waves and promoting unidirectivity. Some of the performance curves obtained on these devices are reproduced.

## Antenna Dimensions

Before designing any highly directive receiving antenna system, it is desirable to have a knowledge of the direction of arrival, the angular spread, and the angular variation in direction of the waves to be

* Published in *Proc. I. R. E.*, January, 1935.

[1] "Developments in Short-Wave Directive Antennas," by E. Bruce, *Proc. I. R. E.*, August, 1931; *Bell Sys. Tech. Jour.*, October, 1931.

received. Such information has been obtained by employing the methods described by Friis, Feldman and Sharpless.[2]

In general, the principal axis of the antenna departs very little from the true bearing of the station to be received. However, the average vertical direction of the waves in the incident plane and their angular variation have an effect upon the determination of the correct dimensions of a horizontal rhombic receiving antenna.

Without careful consideration, one might be led to believe that the theoretical effectiveness of a horizontal rhombic receiving antenna would increase without limit, for a stable wave-direction, as the properly related dimensions are increased. It will be shown that, for a given incident wave-angle above the horizontal, the optimum dimensions have quite definite values.

The essential dimensions of this type of antenna are its height above ground, the length of a side element, and the inclination angle of these elements in respect to the wave-direction. At frequencies higher than about fifteen megacycles, this type of antenna is relatively so inexpensive that it is usually economically possible to select a combination of these dimensions which will produce the maximum signal output.

At lower frequencies, the increasing cost, particularly of the pole height required for maximum output, demands that a careful analysis be made of effective compromise designs. One such design method involves a sacrifice in antenna height, which sacrifice can be partially compensated for by an increase in element length. Thus the economic balance involves a weighing of supporting structure costs against the cost of land. Again, cases may arise where the recommended antenna height is practical but a sacrifice in element length is essential. The effect of this sacrifice on the directive pattern can be partially compensated for by a readjustment of the tilt-angle of the elements. The directive pattern may also be aligned with the wave-angle when both height and element length are reduced, but only at an output sacrifice.

The designer's first problem is to choose between the maximum output arrangement and the various compromise designs. In the case of a transmitting antenna, the problem is to create as large a field as possible at the distant receiving point, for a given amount of transmitter power. As long as the increase in antenna cost is smaller than the corresponding cost of an increase in transmitter power, no compromise should be made in the antenna dimensions. It is, therefore, obvious that the choice, between optimum and compromise dimensions for a transmitting antenna, is a function of the cost of the transmitter to be employed.

[2] "The Determination of the Direction of Arrival of Short Radio Waves," by H. T. Friis, C. B. Feldman and W. M. Sharpless, *Proc. I. R. E.*, January, 1934.

For a receiving antenna, the dividing line between the optimum and the compromise design methods is often defined by the relative importance, as circuit limiting factors, of static as compared with set-noise. The optimum design results in a large signal output, from the antenna, which enables the over-riding of set-noise and it also results in effective directional discrimination against static. The compromise height arrangement maintains and sometimes improves upon the directional discrimination against static but sacrifices a part of the possible antenna output-signal level. If the frequency is sufficiently low so that static rather than set-noise is practically always the circuit limitation, no real harm will result from using the compromise height design with its accompanying economic saving.

In general, compromise in element length results in a loss in both the antenna output-signal level and in static discrimination. This procedure is recommended only for exceptional cases where a restriction in available land exists, or where a broad directional characteristic is required because of a highly variable wave-direction.

## Directivity Equations

In the appendix of this paper will be found the derivation, for the stated assumptions, of the three dimensional directivity equation for a horizontal rhombic receiving antenna terminated by its characteristic impedance. The principal antenna dimensions will be apparent by reference to Fig. 1.

If the horizontal component $\beta$, of the angle made by the wave-direction with the principal antenna axis, is set equal to zero in the final resulting equation in the appendix, the directivity equation for the incident plane passing through the principal axis is as follows:

$$I_R = k\left[1 + ae^{-j\left(\frac{4\pi}{\lambda}H\sin\Delta+\alpha\right)}\right] \cdot \left[\frac{\cos\phi}{1-\sin\phi\cdot\cos\Delta}\right]\cdot\left[1-e^{-j\frac{2\pi l}{\lambda}(1-\sin\phi\cdot\cos\Delta)}\right]^2, \quad (1)$$

where $I_R$ = receiver current,
$k$ = proportionality factor,
$H$ = height above ground,
$l$ = wire element length,
$\phi$ = one-half of side apex angle,
$\Delta$ = angle made with ground by wave-direction in incident plane,
$a$ = amplitude ratio of ground reflected field to incident field,
$\alpha$ = apparent phase angle lag caused by ground reflection,
$\lambda$ = wave-length.

Since the principal axis of the antenna is nearly always directed toward the distant transmitter, the conditions which will maximize the above expression are of prime interest. The determination of these conditions is greatly simplified by the assumption of a perfect ground.

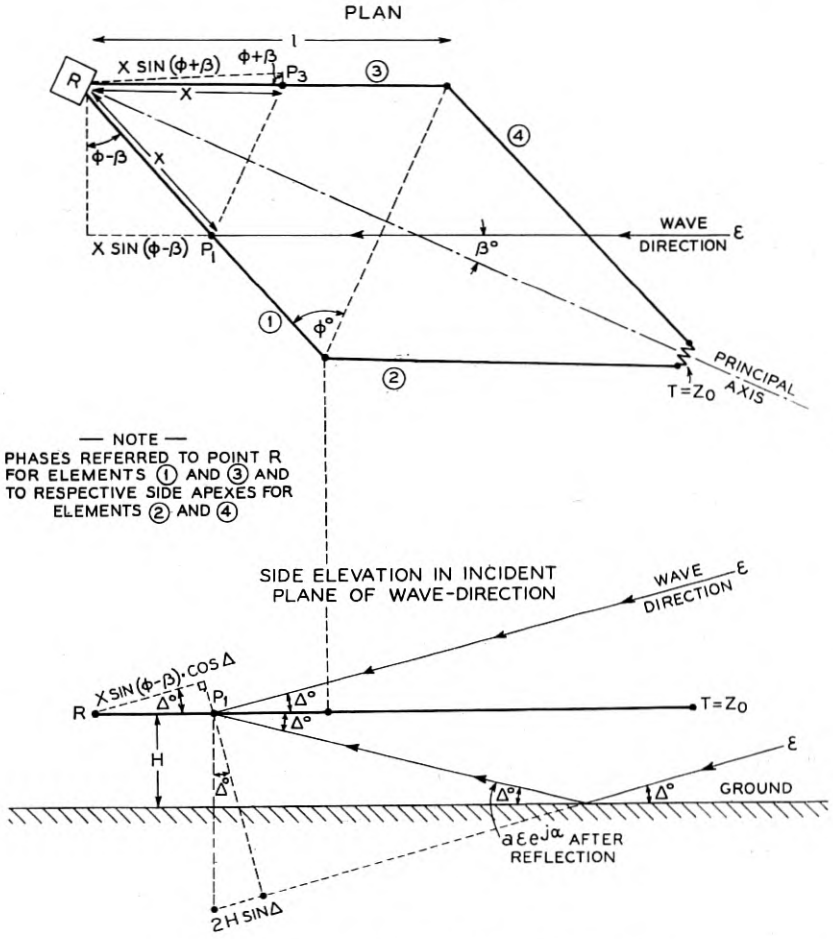


Fig. 1—Horizontal rhombic antenna dimensions.

Fortunately this is not a radical assumption for the case of horizontally polarized waves at the usual incident angles. This may be verified by reference to Fig. 2 where the values of $a$ and $\alpha$ are plotted for several ordinarily encountered ground constants.

Using the perfect ground assumption, the amplitude of equation (1) is given by

$$I_R = k' \left[ \sin \left( \frac{2\pi H}{\lambda} \sin \Delta \right) \right] \cdot \left[ \frac{\cos \phi}{1 - \sin \phi \cdot \cos \Delta} \right] \cdot \left\{ \sin^2 \left[ \frac{\pi l}{\lambda} (1 - \sin \phi \cdot \cos \Delta) \right] \right\}. \quad (2)$$

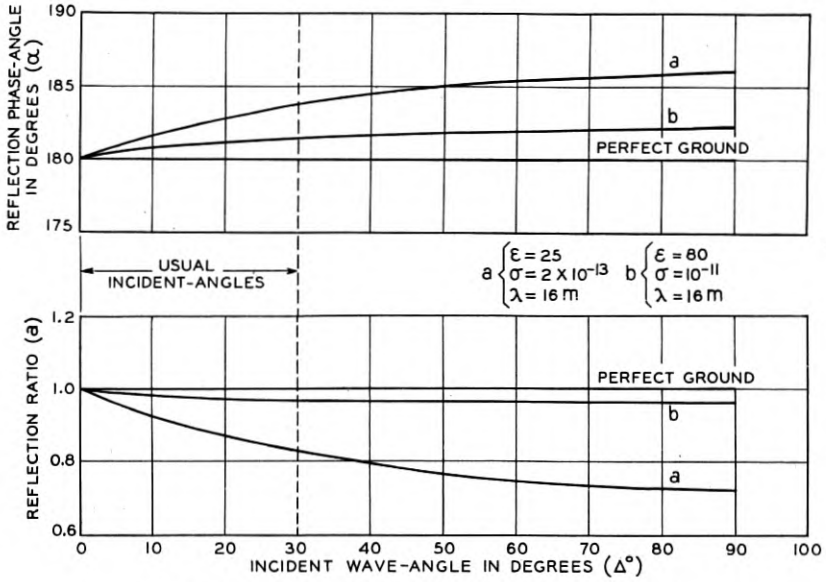


Fig. 2—Characteristics of wave-reflections from ground.

In this equation the first bracketed term may be referred to as the "height" factor, the second as the "asymmetrical directivity" factor, and the third as the "phasing" factor.

## Maximum Output Design Method

In maximizing equation (2), it is necessary to deal with three variables, the antenna dimensions $H$, $l$ and $\phi$. Differentiating with respect to each of these variables, while holding the other two constant, and equating to zero, the following three expressions are obtained:

$\frac{\delta I_R}{\delta H} = 0$ when

$H = \frac{\lambda}{4 \sin \Delta}$ for the lowest practical height,

$\frac{\delta I_R}{\delta l} = 0$ when

$$l = \frac{\lambda}{2(1 - \sin\phi \cdot \cos\Delta)} \quad \text{for the major ear of the directive diagram,} \tag{4}$$

$\frac{\delta I_R}{\delta \phi} = 0$ when

$$\tan\left[\frac{\pi l}{\lambda}(1 - \sin\phi \cdot \cos\Delta)\right] = -\frac{2\pi l \cdot \cos^2\phi \cdot \cos\Delta(1 - \sin\phi \cdot \cos\Delta)}{\lambda(\sin\phi - \cos\Delta)}. \tag{5}$$

Substituting equation (4) into (5), it is found that,

$$\sin\phi = \cos\Delta. \tag{6}$$

This result determines the fact that, regardless of antenna height, the best tilt-angle $\phi$ is the complement of the wave-angle $\Delta$, where the optimum element length is used. Taking this result and substituting it back into either (4) or (5) results in

$$l = \frac{\lambda}{2\sin^2\Delta}. \tag{7}$$

The value of "$l$" in equation (7) together with the height given by equation (3), and the tilt-angle given by equation (6) determine the dimensions of the horizontal rhombic antenna with maximum output for any given wave-angle $\Delta$.

As an example of the use of the above equations, Fig. 3 is the resulting incident plane directive pattern with the antenna designed for the given wave-angle of 17.5 degrees. Figure 4 is a plot of the antenna output versus the azimuth angle for the fixed incident angle of 17.5 degrees as obtained by using these dimensions in the three-dimensional directivity equation in the appendix.

Note in Fig. 3 that the directive pattern does not have its maximum radius at the line indicating the given wave-direction even though the greatest possible amplitude for that wave-angle has been determined.

It is believed that this method of design has a definite application where over-riding set-noise by the largest possible signal output is paramount. In cases where discrimination against random static is desirable, or where the received wave-direction is unstable, an alignment of the mean wave-direction with the optimum radius of the

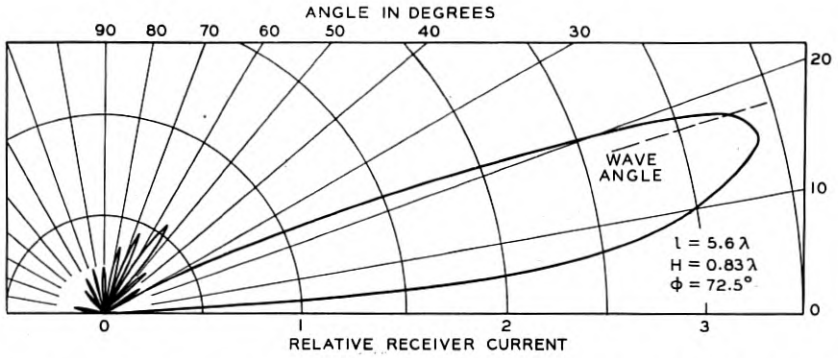


Fig. 3—Maximum output design. Receiver current diagram of incident plane directivity for a 17.5° wave-angle.

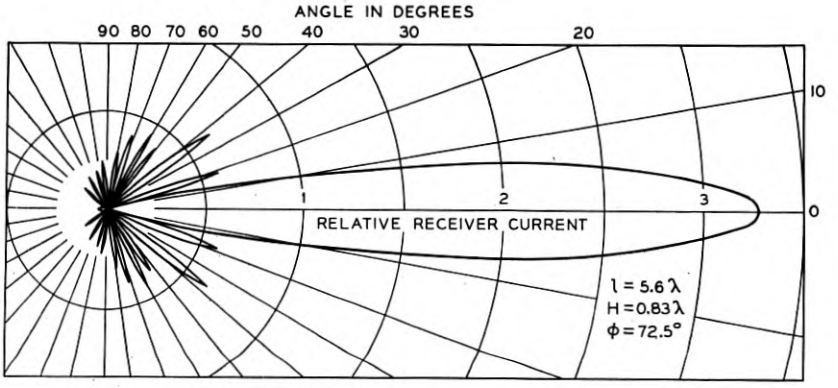


Fig. 4—Maximum output design. Receiver current versus azimuth angle for a 17.5° wave-angle.

directive diagram becomes necessary. Such an alignment can be achieved by the method described below, but only by a small sacrifice in the amplitude [3] of $I_R$ at the required wave-angle.

## The Alignment Design Method

Since by our previous method of design the height factor of equation (2) was aligned with the wave-angle, it is necessary to change either or both of the dimensions $l$ and $\phi$ to secure the best output under the limiting alignment condition. This limiting condition is obtained as

[3] To compare the energy gain of one antenna with another, the amplitudes of $I_R$, which are proportional to effective voltages, should be corrected for the differences in radiation resistance of the antennas.

follows:

$$\frac{\delta I_R}{\delta \Delta} = 0 \text{ when}$$

$$H = \sin\phi \cdot \tan\Delta \cdot \tan\left(\frac{2\pi H}{\lambda}\sin\Delta\right) \cdot \left\{\frac{\lambda}{2\pi(1 - \sin\phi\cdot\cos\Delta)} - \frac{l}{\tan\left[\frac{\pi l}{\lambda}(1 - \sin\phi\cdot\cos\Delta)\right]}\right\}. \quad (8)$$

Substituting (3) into (8) gives

$$\tan\left[\frac{\pi l}{\lambda}(1 - \sin\phi\cdot\cos\Delta)\right] = 2\left[\frac{\pi l}{\lambda}(1 - \sin\phi\cdot\cos\Delta)\right]. \quad (9)$$

This is a transcendental equation of the form $\tan x = 2x$, whose roots, to four significant figures, are

$$x = 0, \quad 0.3710\pi, \quad 1.466\pi, \quad 2.480\pi, \quad 3.486\pi, \quad \text{etc.}$$

Alignment of the major ear of the directive pattern with the wave-angle occurs only for the first solution greater than zero. Therefore we have for this condition,

$$l = \frac{0.371\lambda}{1 - \sin\phi\cdot\cos\Delta}. \quad (10)$$

To obtain the dimensions for greatest output in the alignment design case, $l$ and $\phi$ in equation (2) are no longer independent variables, but are related by equation (10). We can, therefore, eliminate either one we choose from equation (2) and maximize this new equation for the required dimensions.

Substituting (10) into (2) gives:

$$I_R' = k''\left[\sin\left(\frac{2\pi H}{\lambda}\sin\Delta\right)\right]\left[\frac{\cos\phi}{1 - \sin\phi\cos\Delta}\right] \quad (11)$$

and

$$\frac{\delta I_R'}{\delta\phi} = 0 \text{ when} \quad (12)$$

$$\sin\phi = \cos\Delta.$$

Since equation (12) is identical with equation (6), the recommended design for directive pattern alignment with the wave-angle is obtained by changing only the antenna length which was given for the maximum

output design by equation (7). By substituting (12) into (10), this becomes

$$l = \frac{0.371\lambda}{\sin^2 \Delta} . \tag{13}$$

Therefore, alignment design is obtained by the use of equations (3), (6) and (13). This shows that to change from maximum output design to alignment design it is only necessary to reduce the length to approximately seventy-four per cent of the value recommended for maximum output.

Figures 5 and 6 are examples of the directive patterns secured by such

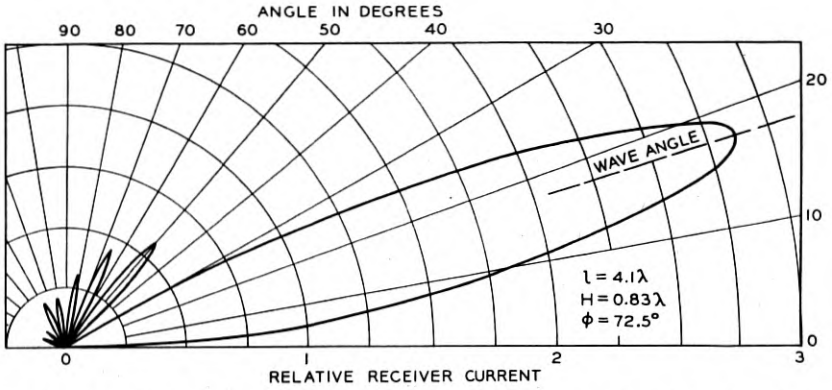


Fig. 5—Alignment design. Receiver current diagram of incident plane directivity for a 17.5° wave-angle.

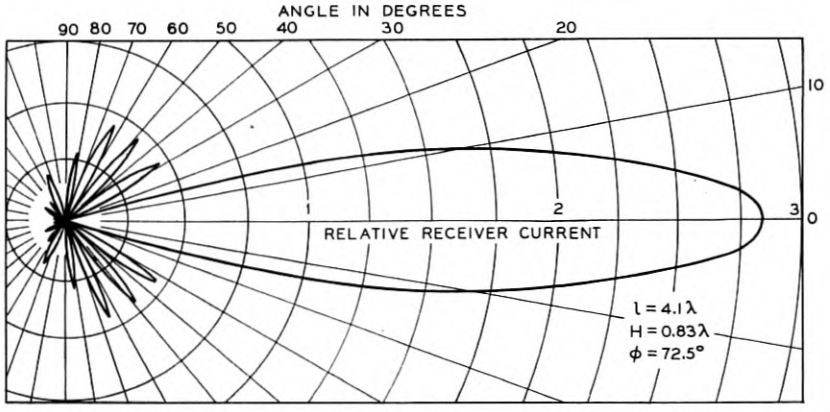


Fig. 6—Alignment design. Receiver current versus azimuth angle for a 17.5° wave-angle.

alignment design for a wave-angle of 17.5 degrees. Comparison of these diagrams with Figs. 3 and 4 shows a small loss in $I_R$ resulting from such design. Figures 7 and 8 are similar alignment design plots for a wave-angle of 11 degrees.

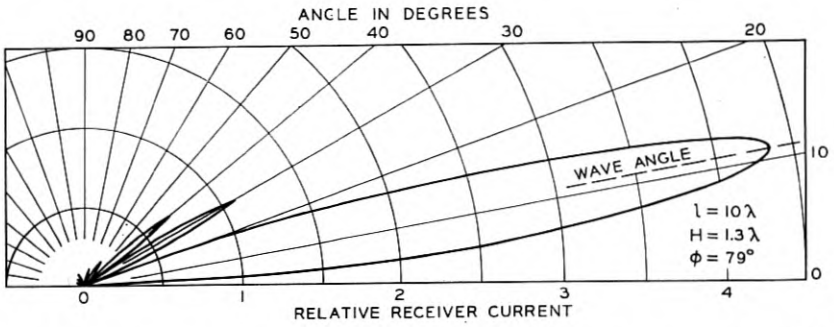


Fig. 7—Alignment design. Receiver current diagram of incident plane directivity for an 11° wave-angle.

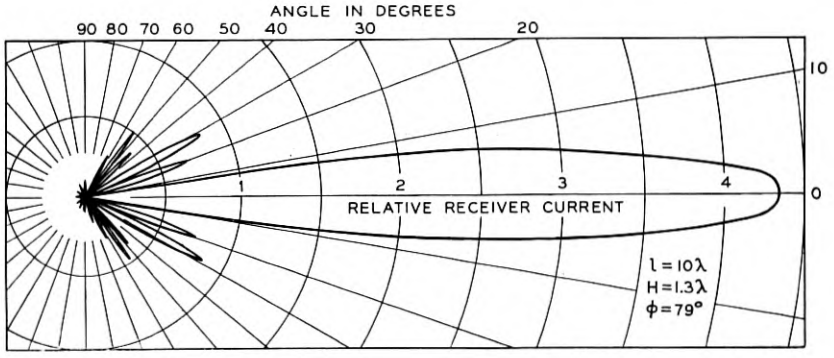


Fig. 8—Alignment design. Receiver current versus azimuth angle for an 11° wave-angle.

## Compromise Design Methods

When an antenna output sacrifice must be tolerated to obtain smaller dimensions, it is often desirable to design the system for directive pattern alignment with the wave-angle, as mentioned above. The methods of accomplishing this result will now be given.

The effect on the directive pattern of reducing the height can be largely compensated for by increasing the length.

Substituting (6) into (8) gives

$$\frac{H}{\tan\left(\frac{2\pi H}{\lambda}\sin\Delta\right)} = \frac{\lambda}{2\pi\sin\Delta} - \frac{l\cdot\sin\Delta}{\tan\left(\frac{\pi l}{\lambda}\sin^2\Delta\right)}. \qquad (14)$$

When both $\Delta$ and $H$ are specified, this equation gives the element length $l$ to be used with the reduced height. Figures 9 and 10 are

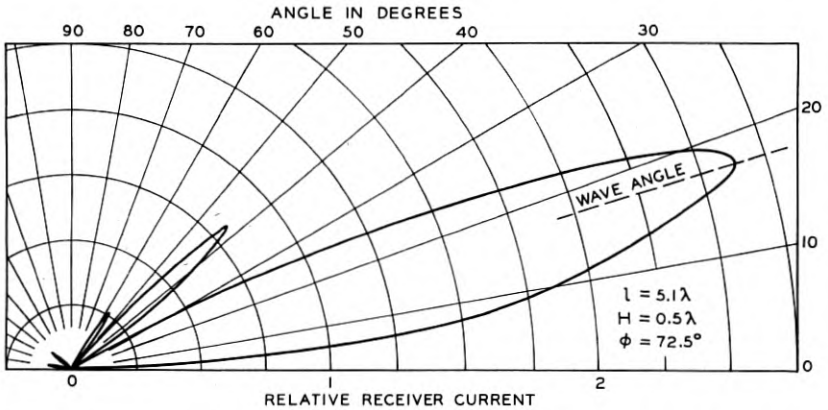


Fig. 9—Compromise height design. Receiver current diagram of incident plane directivity for a 17.5° wave-angle and a height of one-half wave-length.

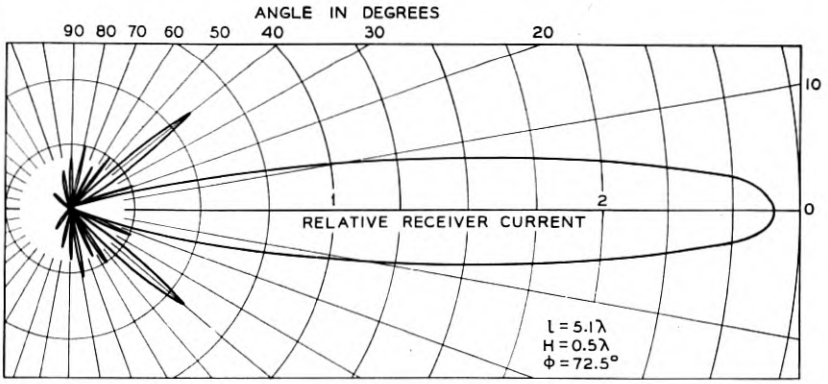


Fig. 10—Compromise height design. Receiver current versus azimuth angle for a 17.5° wave-angle and a height of one-half wave-length.

examples of the directive diagrams resulting from a reduced height of one-half wave-length and an incident wave-angle of 17.5 degrees. These last mentioned figures may be directly compared with Figs. 5

and 6 to determine what effect the sacrifice in height and the increased length has had.

The effect on the directive pattern of reducing the length can be minimized by changing the tilt-angle $\phi$. When the height given by equation (3) is used, the new tilt-angle can be obtained by writing equation (10) as

$$\sin \phi = \frac{l - 0.371\lambda}{l \cos \Delta} . \tag{15}$$

If it is desired to align the directive pattern with the wave-angle when the height and length are both reduced from the recommended values, equation (8) can be solved for the new tilt-angle. Any of these compromise designs will, of course, reduce the antenna output.

## Limitations of Results

The accuracy of the calculations in this paper is based on the following assumptions:

1. Negligible effect by wire attenuation.
2. Negligible mutual coupling between elements.
3. The rate of change of the effective voltage with varying dimensions is large as compared with the rate of change of radiation resistance.

The validity of the assumptions has been checked to engineering accuracies in the ultra-short wave experiments to be described, where the shape of the directivity predicted by these equations agreed well with experiment. However, theory and experiment have both shown that ordinary ground has an appreciable influence on the wire attenuation when the antenna heights are of the order of only small fractions of a wave-length. This effect increases as the antenna dimensions are extended. In such unusual cases, this behavior should be carefully investigated even though calculations including large wire attenuations have shown surprisingly little effect upon the directive diagrams. Approximate values for the variation of radiation resistance with dimensions, for terminated rhombic antennas, indicate that assumption 3 is probably warranted provided that the antenna is not extraordinarily small.

Finally, attention should be called to these important facts:

The suggested dimensions in this paper not only do not restrict the broad frequency range of this type of antenna but are in many cases

the broadest arrangements under the prescribed conditions. This statement is based on a large number of calculated directive diagrams, for various frequencies, and has been substantiated by operating tests.

Also, in certain applications, as with ultra-short waves where the direction of wave propagation is substantially horizontal ($\Delta = 0$), the derived equations place no restrictions on the antenna size.

## Experimental Checks of Directivity Calculations

One method employed for experimentally checking the calculations of directivity was by means of a model small enough to permit continuous rotation of the antenna in respect to a fixed wave-direction. A simple transmitter and a double-detection receiving set, which employed a calibrated intermediate frequency attenuator, both operating at about four meters, were utilized for this work. Even at these wave-lengths, the physical dimensions of the antenna were rather large so that it was necessary to build an antenna smaller in wave-length dimensions than is usually recommended for commercial installations. As previously stated, any antenna for regular service would be designed for its own specific conditions, so that this antenna is not to be taken as a recommended design. However, this fact had no bearing on the results of these tests, for there was no intention of using this movable antenna for any purpose other than experimentally checking the directivity equations.

A photograph of the test installation is shown in Fig. 11. The antenna wires were mounted on a lattice work cross of light weight wood suspended by a system of ropes from two wooden poles. This permitted the hoisting of the entire antenna system to any height up to several wave-lengths, and rotating it to any desired angular position. The antenna is shown in more detail in the photograph of Fig. 12. It was constructed of two space-tapered wires in parallel, spaced in order to lower the antenna impedance to permit the matching of the open-wire transmission line which was attached to one end. Each element was three and a quarter wave-lengths long. The receiving set was placed in a small push-cart at the other end of the transmission line so that the line length did not need to be changed with antenna movements. An antenna of this size had a measured gain of fourteen decibels over a half-wave horizontal dipole at the same height.

The horizontal plane directive pattern was experimentally determined by measuring the antenna angle and signal output at each maximum and minimum and at fixed angles on the major lobe, as the antenna was rotated about its vertical axis. The pattern measured

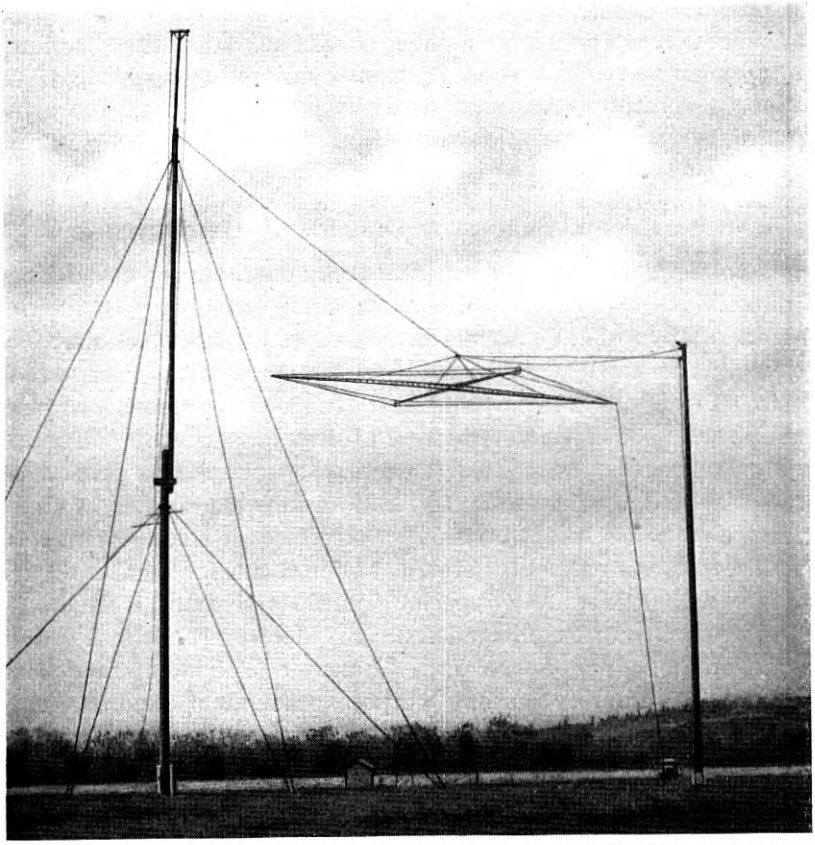

Fig. 11—Test antenna used at four meters for experimental checks of directivity calculations.

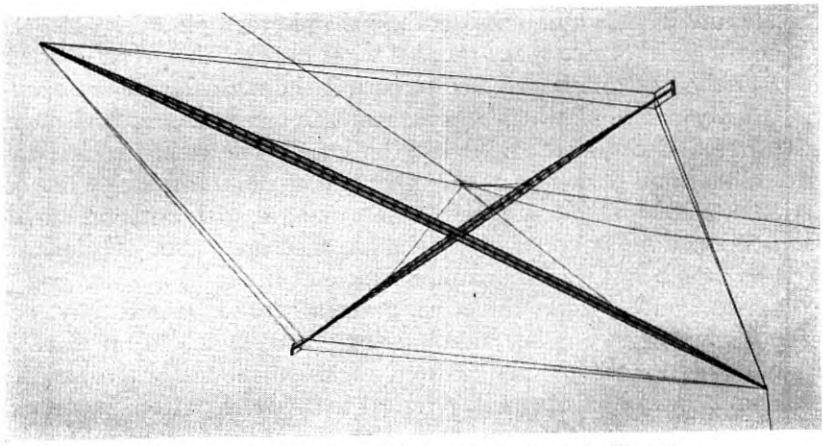

Fig. 12—A detail view of the antenna shown in Fig. 11.

in this manner is shown in Fig. 13. In this figure is also shown, for comparison purposes, the pattern calculated for these conditions by the equations derived in the appendix. The agreement between these two patterns is quite evident.

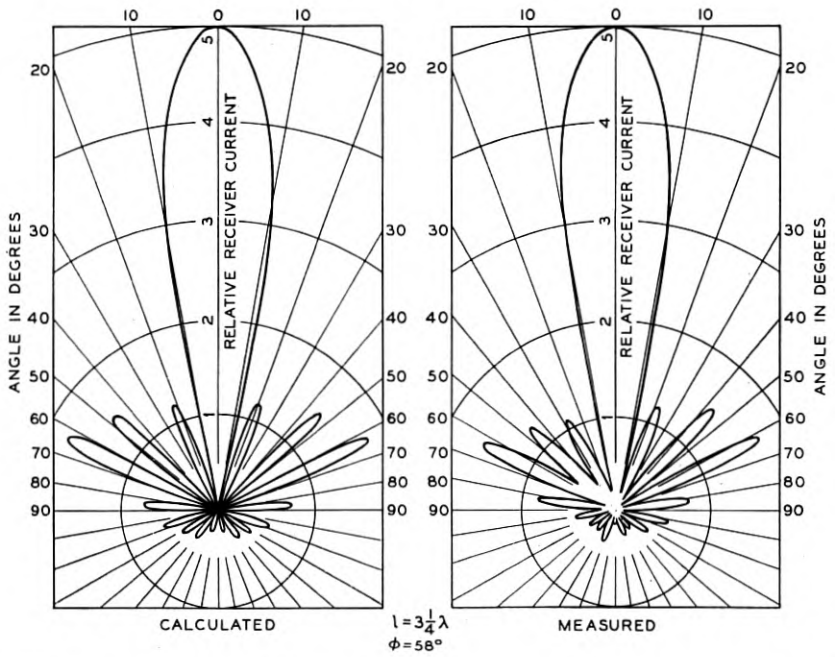


Fig. 13—Receiver current diagrams of test antenna output current versus azimuth angle for a 2.7° wave-angle.

To check the vertical directivity of the system, a portable oscillator was raised from the ground to the top of a one-hundred foot pole in front of the antenna. Due to the limited range of arrival angles so obtained, only the major lobe of the directive pattern could be traced out. Figure 14 gives the calculated major lobe for these conditions (equations corrected to actual path length difference between direct and reflected waves) with the antenna one-half wave-length above ground, and Fig. 15 is the pattern at a height of one wave-length. The circles on these curves are the measured values. These patterns show that the directivity does not change very rapidly with changes in antenna height.

Experimental checks of these directivity calculations have also been obtained on the larger antennas ordinarily used for long distance reception. A large horizontal rhombic antenna was constructed with a

motor driven winch to adjust its interior angles, thus providing control of the directivity of the antenna. This antenna could be "steered" so that either the major lobe or the first null could be used to receive or

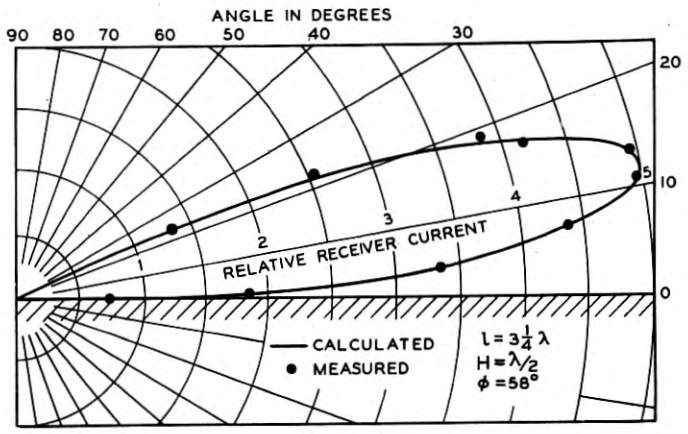


Fig. 14—Receiver current diagram of incident plane directivity with test antenna one-half wave-length above ground. Major lobe only.

discriminate against the components of the received signal. The angle of arrival of these components was measured by the methods described

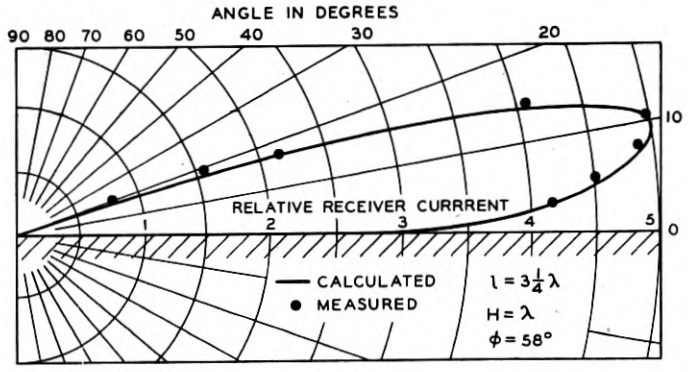


Fig. 15—Receiver current diagram of incident plane directivity with test antenna one wave-length above ground. Major lobe only.

by Friis, Feldman and Sharpless.[2] The agreements were at all times close. This work was incidental to some fading studies the results of which also checked qualitatively the directivity calculations.

These substantial agreements between calculated and measured values are an indication of the accuracy secured in theoretically predicting antenna directivity when using the stated assumptions. It is believed, therefore, that they establish the validity of the design methods now being used for commercial applications of this type of antenna.

## Standing Wave Suppression

To obtain the uni-directional characteristics usually desired and to provide a constant terminal impedance over a broad frequency range, the end of the rhombic antenna, remote from the receiver, is terminated in its equivalent characteristic impedance to prevent wave reflections.

The termination was originally adjusted by determining the greatest front-to-back signal reception ratios which could be obtained with the terminating graphite resistors or the other circuits employed. This method of testing proved both cumbersome and erratic for the following reasons:

Radiating field oscillators had to be placed at sizable distances both in front and in the rear of the antenna. Manipulating these oscillators, for frequency runs, proved inconvenient. Furthermore, properly elevating the oscillators was impractical because of the large heights required to simulate the actual average wave-angles usually encountered. When the oscillators were at more reasonable heights above ground, the front-to-back signal ratios could rarely be reproduced since the wave-direction was unstably situated on the lower steep edge of the incident plane directive diagram.

The procedure for testing the termination has been simplified through the measurement, by a substitution method, of the antenna terminal impedance as the frequency is varied over the required frequency band. By readjusting the termination, until these measured values are constant, the required value is obtained.

The schematic diagram of the impedance measuring equipment employed is shown in Fig. 16. In the actual apparatus, the individual

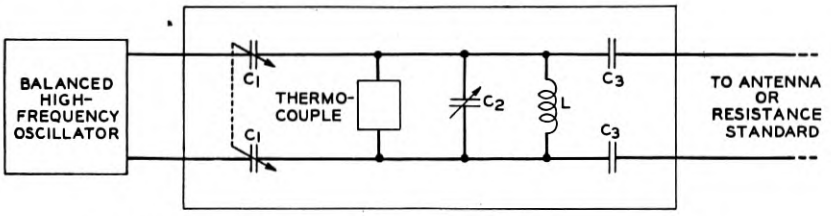


Fig. 16—Schematic diagram of impedance measuring equipment.

parts are separately shielded and symmetrically arranged so that the assembly can be attached to a portable balanced oscillator used for other work. Such a combination is shown in the photograph of Fig. 17.

Again referring to Fig. 16, the condensers marked $C_1$ control the degree of coupling while maintaining an accurate balance to ground.

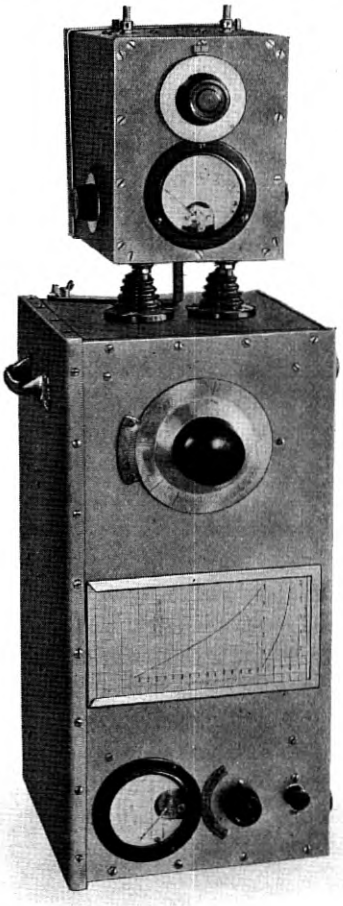

Fig. 17—Impedance measuring equipment and balanced oscillator.

$C_2$ is a condenser used for anti-resonating the complete circuit, this anti-resonance being indicated by a thermocouple deflection. The condensers labelled $C_3$ are large blocking condensers which have been added to the circuit to permit the measurement of the value of the variable substitution resistance without the necessity of disconnecting that resistance. A pencil lead is usually employed as this resistor.

The measurement consists in replacing the antenna by a value of resistance $R$ which gives the same thermocouple deflection. The reactive components involved are tuned out in each case by adjusting the condenser $C_2$. The change in the dial setting of this condenser, as the resistance is substituted for the antenna, permits the determination of the reactive component of the antenna impedance.

Curve 3, in Fig. 18, shows the terminal resistance versus frequency of a properly terminated experimental antenna constructed by the authors. Curve 1, of that figure, results from a termination resistance which is too large while that in curve 2 is too small. Incidentally, it is interesting to note that the value of the measured terminal resistance of this antenna is lower than the resistance required, at the other end of the antenna, for the correct termination. This behavior is probably an effect of the radiation resistance of the system.

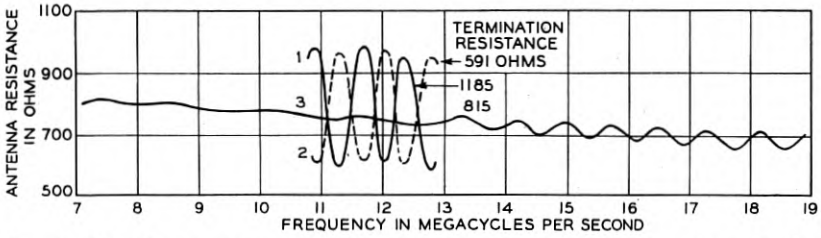


Fig. 18—Rhombic antenna resistance versus frequency for three values of termination resistance.

When a lumped termination resistance is employed, measurements reveal that it operates as if paralleled by a small effective capacitance, a capacitance probably due to the proximity of the adjacent converging antenna wires. This capacitance prevents the resistance from acting as an entirely satisfactory termination. Its effect can be minimized through a proper distribution of the termination resistance over a short length of the converging part of the antenna. This distribution tends to isolate the harmful capacitances by means of the resistances and also introduces a small effective series inductance which is helpful in neutralizing the residual capacitive reactance. Other circuits have been devised, but this one is often employed because of its simplicity and ruggedness.

### Antenna-to-Pipe Line Coupling Circuits

Concentric pipe transmission lines are usually favored for receiving antenna installations as they can be buried in the ground to give a substantial and weather-proof construction, free from reactions on adjacent antennas. This pipe-line runs up one pole of a rhombic

antenna and terminates adjacent to the output terminals of the antenna.

Large terminated rhombic antennas, built with number 12 A.W.G. wire, present a terminal impedance, balanced to ground, of nearly 800 ohms resistance. Where the ratio of the inner diameter of the outer pipe to the outer diameter of inner pipe of the transmission line is the optimum value [5] of 3.6, the characteristic impedance of the line is about 77 ohms. The problem is, therefore, to transform properly a balanced resistance of about 800 ohms into an unbalanced resistance of about 80 ohms without appreciable losses over the entire operating range of the antenna.

Figure 19 is a photograph of a simple type of coupling circuit developed for experimental use. Figure 20 is the measured loss versus frequency curve obtained when the device is connected between the stated impedances. The broad frequency range is primarily the result of a high coefficient of coupling. Balance is obtained by symmetrically disposing two primary windings in series about two secondary windings in parallel. A still greater operating range may be obtained through a more elaborate arrangement. Note in Fig. 19 that the boxes and pipe line are gas tight so that nitrogen pressure can be applied, if desired, to prevent moisture absorption through "breathing."

## Conclusion

The original paper [1] on rhombic antennas, in discussing the situation regarding short-wave communication, stated that the main limitations in this field were three in number:

(*a*) Inherent receiver noise.
(*b*) External noise (static, man-made noises, etc.)
(*c*) Signal fading.

That paper attempted to show that the design of the receiving antenna system has an important bearing upon overcoming all three of these difficulties.

In the present paper, the authors have indicated how progress in overcoming the first two limitations, through refinements in the horizontal rhombic type of antenna, has been made.

Improvements in the selection of the receiving antenna dimensions have greatly helped the noise situation. Termination and coupling circuit developments now allow many of these antennas to operate one circuit or several circuits simultaneously within a four-to-one frequency range.

[5] E. J. Sterba and C. B. Feldman, "Transmission Lines for Short-Wave Radio Systems," *Proc. I. R. E.*, July, 1932; *Bell Sys. Tech. Jour.*, July, 1932.

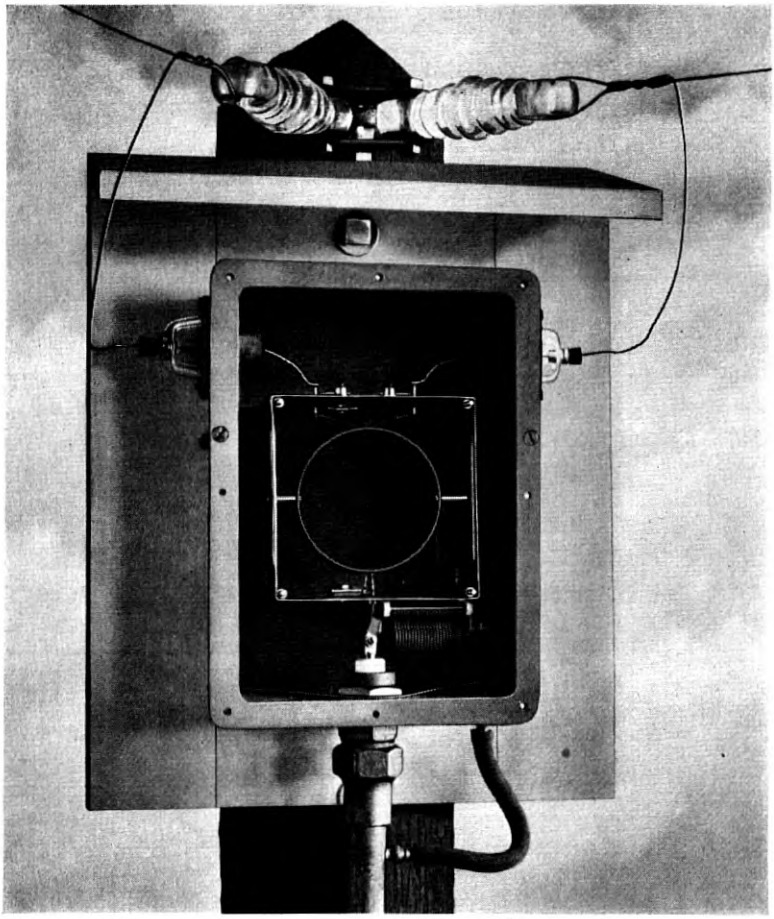

Fig. 19—Experimental equipment for coupling a rhombic antenna to a concentric pipe transmission line.

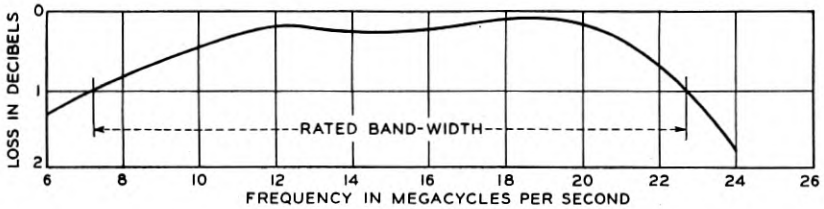


Fig. 20—Loss versus frequency of the experimental coupling equipment, shown in Fig. 19, when used to connect an antenna balanced impedance of 778 ohms in parallel with 5 mmf. to a pipe-line unbalanced resistance of 80 ohms.

## APPENDIX

### Calculations of Three-Dimensional Directivity

Antenna: Horizontal Rhombus.
Wave Polarization: Horizontal.
Far End Termination: Characteristic Impedance.
Ground Constants: General Case.

Assumptions: Negligible Wire Attenuation and Leakage.
Negligible Mutual Coupling.
Uniform Characteristic Impedance.

Nomenclature: (See Fig. 1)

$P$ = point under consideration.

$l$ = element length.

$\phi$ = one-half of side apex angle.

$\Delta$ = angle made with ground by wave-direction in incident plane.

$\beta$ = angle made by horizontal component of wave-direction with principal axis drawn through receiver and termination.

$H$ = height above ground.

$X$ = wire distance of point $P$ from $R$ for element (2) as well as for element (1).

$\epsilon$ = r.m.s. space voltage.

$E$ = r.m.s. space voltage and phase at $P$ due to $\epsilon$, direct propagation.

$E'$ = r.m.s. space voltage and phase at $P$ due to $\epsilon$, after ground reflection.

$E_{P1}$ and $E_{P2}$ = wire voltage at $P$ in element (1) or element (2), direct propagation.

$E_{P1}'$ and $E_{P2}'$ = same as above but for ground reflection.

$a$ = amplitude ratio of ground reflection to direct space voltage.

$\alpha$ = apparent phase angle change after ground reflection.

$\lambda$ = wave-length.

$I_R$ = receiver current.

$Z_0$ = characteristic impedance.

$R$ = receiver impedance.

$T$ = far end termination impedance.

Voltages Induced in Wire:

For a directly propagated wave in element (1), $R$ being the reference point for phases,

$$E_{P1} = \epsilon \cdot \cos(\phi - \beta) \cdot e^{+j\frac{2\pi}{\lambda} \cdot \sin(\phi - \beta) \cdot \cos\Delta \cdot X} dX.$$

For reflected wave in element (1) ($X = 0$ to $l$)

$$E_{P1}' = a \cdot \epsilon \cdot \cos(\phi - \beta) \cdot e^{+j\left\{\frac{2\pi}{\lambda}[X \cdot \sin(\phi - \beta) \cdot \cos\Delta - 2H \sin\Delta] - \alpha\right\}} dX.$$

Similarly for element (3) ($X = 0$ to $l$)

$$E_{P3} = \epsilon \cdot \cos(\phi + \beta) \cdot e^{+j\frac{2\pi}{\lambda} \cdot \sin(\phi + \beta) \cdot \cos\Delta \cdot X} dX.$$

$$E_{P3}' = a \cdot \epsilon \cdot \cos(\phi + \beta) \cdot e^{+j\left\{\frac{2\pi}{\lambda}[X \cdot \sin(\phi + \beta) \cdot \cos\Delta - 2H \sin\Delta] - \alpha\right\}} dX.$$

Also for elements (2) and (4) ($X = 0$ to $l$) where respective side apexes are now reference points for phases,

$$E_{P2} = -E_{P3} \cdot e^{+j\frac{2\pi l}{\lambda} \cdot \sin(\phi - \beta) \cdot \cos\Delta},$$

$$E_{P2}' = -E_{P3}' \cdot e^{+j\frac{2\pi l}{\lambda} \cdot \sin(\phi - \beta) \cdot \cos\Delta},$$

and

$$E_{P4} = -E_{P1} \cdot e^{+j\frac{2\pi l}{\lambda} \cdot \sin(\phi + \beta) \cdot \cos\Delta},$$

$$E_{P4}' = -E_{P1}' \cdot e^{+j\frac{2\pi l}{\lambda} \cdot \sin(\phi + \beta) \cdot \cos\Delta}.$$

Receiver Current where $T = Z_0$ and $R = Z_0$

$$I_R = \int_0^l \frac{E_{P1} + E_{P1}' + E_{P3} + E_{P3}'}{2Z_0} \cdot e^{-j\frac{2\pi}{\lambda}X} + \int_0^l \frac{E_{P2} + E_{P2}' + E_{P4} + E_{P4}'}{2Z_0} \cdot e^{-j\frac{2\pi}{\lambda}X} \cdot e^{-j\frac{2\pi}{\lambda}l},$$

$$I_R = \int_0^l \frac{\epsilon \cdot \cos(\phi - \beta)}{2Z_0} \cdot \left[1 + ae^{-j\left(\frac{4\pi}{\lambda} H \sin\Delta + \alpha\right)}\right] \cdot e^{-j\frac{2\pi}{\lambda}[1 - \sin(\phi - \beta) \cdot \cos\Delta]X} dX + \int_0^l \frac{\epsilon \cdot \cos(\phi + \beta)}{2Z_0} \cdot \left[1 + ae^{-j\left(\frac{4\pi}{\lambda} H \sin\Delta + \alpha\right)}\right] \cdot e^{-j\frac{2\pi}{\lambda}[1 - \sin(\phi + \beta) \cdot \cos\Delta]X} dX - \int_0^l \frac{\epsilon \cdot \cos(\phi - \beta)}{2Z_0} \cdot \left[1 + ae^{-j\left(\frac{4\pi}{\lambda} H \sin\Delta + \alpha\right)}\right] \cdot e^{-j\frac{2\pi l}{\lambda}[1 - \sin(\phi + \beta)\cos\Delta]} \cdot e^{-j\frac{2\pi}{\lambda}[1 - \sin(\phi - \beta)\cos\Delta]X} dX$$

$$- \int_0^l \frac{\epsilon \cdot \cos(\phi + \beta)}{2Z_0} \cdot \left[1 + ae^{-j\left(\frac{4\pi}{\lambda} H \sin\Delta + \alpha\right)}\right] \cdot e^{-j\frac{2\pi l}{\lambda}[1 - \sin(\phi - \beta)\cos\Delta]} \cdot e^{-j\frac{2\pi}{\lambda}[1 - \sin(\phi + \beta)\cos\Delta]X} dX,$$

$$I_R = \frac{\epsilon \cdot \cos(\phi - \beta)}{2Z_0} \cdot \left[1 + ae^{-j\left(\frac{4\pi}{\lambda} H \sin\Delta + \alpha\right)}\right] \cdot \left[1 - e^{-j\frac{2\pi l}{\lambda}[1 - \sin(\phi + \beta)\cdot\cos\Delta]}\right] \cdot \int_0^l e^{-j\frac{2\pi}{\lambda}[1 - \sin(\phi - \beta)\cdot\cos\Delta]X} dX$$
$$+ \frac{\epsilon \cdot \cos(\phi + \beta)}{2Z_0} \cdot \left[1 + ae^{-j\left(\frac{4\pi}{\lambda} H \sin\Delta + \alpha\right)}\right] \cdot \left[1 - e^{-j\frac{2\pi l}{\lambda}[1 - \sin(\phi - \beta)\cdot\cos\Delta]}\right] \cdot \int_0^l e^{-j\frac{2\pi}{\lambda}[1 - \sin(\phi + \beta)\cdot\cos\Delta]X} dX,$$

$$I_R = +j\frac{\epsilon\lambda}{4\pi Z_0} \cdot \frac{\cos(\phi - \beta)}{1 - \sin(\phi - \beta)\cdot\cos\Delta} \cdot \left[1 + ae^{-j\left(\frac{4\pi}{\lambda} H \sin\Delta + \alpha\right)}\right] \cdot \left[1 - e^{-j\frac{2\pi l}{\lambda}[1 - \sin(\phi + \beta)\cdot\cos\Delta]}\right]\left[1 - e^{-j\frac{2\pi l}{\lambda}[1 - \sin(\phi - \beta)\cdot\cos\Delta]}\right]$$
$$+ j\frac{\epsilon\lambda}{4\pi Z_0} \cdot \frac{\cos(\phi + \beta)}{1 - \sin(\phi + \beta)\cdot\cos\Delta} \cdot \left[1 + ae^{-j\left(\frac{4\pi}{\lambda} H \sin\Delta + \alpha\right)}\right] \cdot \left[1 - e^{-j\frac{2\pi l}{\lambda}[1 - \sin(\phi - \beta)\cdot\cos\Delta]}\right]\left[1 - e^{-j\frac{2\pi l}{\lambda}[1 - \sin(\phi + \beta)\cdot\cos\Delta]}\right],$$

$$I_R = j\frac{\epsilon\lambda}{4\pi Z_0} \cdot \left[\frac{\cos(\phi - \beta)}{1 - \sin(\phi - \beta)\cdot\cos\Delta} + \frac{\cos(\phi + \beta)}{1 - \sin(\phi + \beta)\cdot\cos\Delta}\right] \cdot \left[1 + ae^{-j\left(\frac{4\pi}{\lambda} H \sin\Delta + \alpha\right)}\right] \cdot \left[1 - e^{-j\frac{2\pi l}{\lambda}[1 - \sin(\phi + \beta)\cdot\cos\Delta]}\right]$$
$$\cdot \left[1 - e^{-j\frac{2\pi l}{\lambda}[1 - \sin(\phi - \beta)\cdot\cos\Delta]}\right],$$

which is the final equation including phase relations.

# Extraneous Frequencies Generated in Air Carrying Intense Sound Waves *

By A. L. THURAS, R. T. JENKINS and H. T. O'NEIL

The exact equation of the propagation of plane sound waves in air is not linear and consequently harmonics and combination tones are generated. The pressure of these extraneous frequencies in terms of the fundamental pressure, frequency, and distance from the source has been mathematically determined by Rayleigh, Lamb and others. These equations have been applied to an exponential horn.

Measurements of the second harmonic and combination tones have been made at various points within a long tube, and in front of an exponential horn. Measurements, in general, agree with theory, but the absolute values are lower than the calculated values.

RECENT developments in horn type loud speakers [1] for high quality reproduction of intense sounds necessitate a consideration of the more exact equations of wave motion if distortion due to the generation of extraneous frequencies in the air of the horn itself is to be avoided. Similar considerations may be of some importance in connection with the pick-up of intense sounds.

The propagation of waves of finite displacement has interested physicists for more than a century. In 1808 Poisson derived an equation which shows that, in general, a sound wave cannot be propagated without a change in form and consequent generation of additional frequencies. This distortion is caused by the non-linearity of air; that is, if equal positive and negative increments of pressure are impressed on a mass of air the changes in volume of the mass will not be equal; the volume change for the positive pressure will be less than the volume change for the equal negative pressure. An idea of the nature of the distortion can be obtained from the adiabatic curve $AB$ for air as given in the familiar volume pressure indicator diagram (Fig. 1a). The undisturbed pressure and specific volume of air are indicated by point $P_0V_0$. Any deviation from the tangent through this point causes distortion and consequent generation of extraneous frequencies. The theoretical magnitudes of the waves of extraneous frequencies are obtained from a solution of the exact differential equation of wave propagation in air. The solution shows that the pressure of the second harmonic frequency, which is generated in the air, increases with the frequency and the magnitude of the fundamental

* Published in the January 1935 issue of the *Jour. Acous. Soc. Am.*

[1] E. C. Wente and A. L. Thuras, "Loud Speakers and Microphones," *Bell Sys. Tech. Jour.*, 13, 259 (1934).

pressure and also with the distance from the sound source. The solution also gives the magnitudes of the waves of sum and difference frequencies generated when two tones are simultaneously impressed

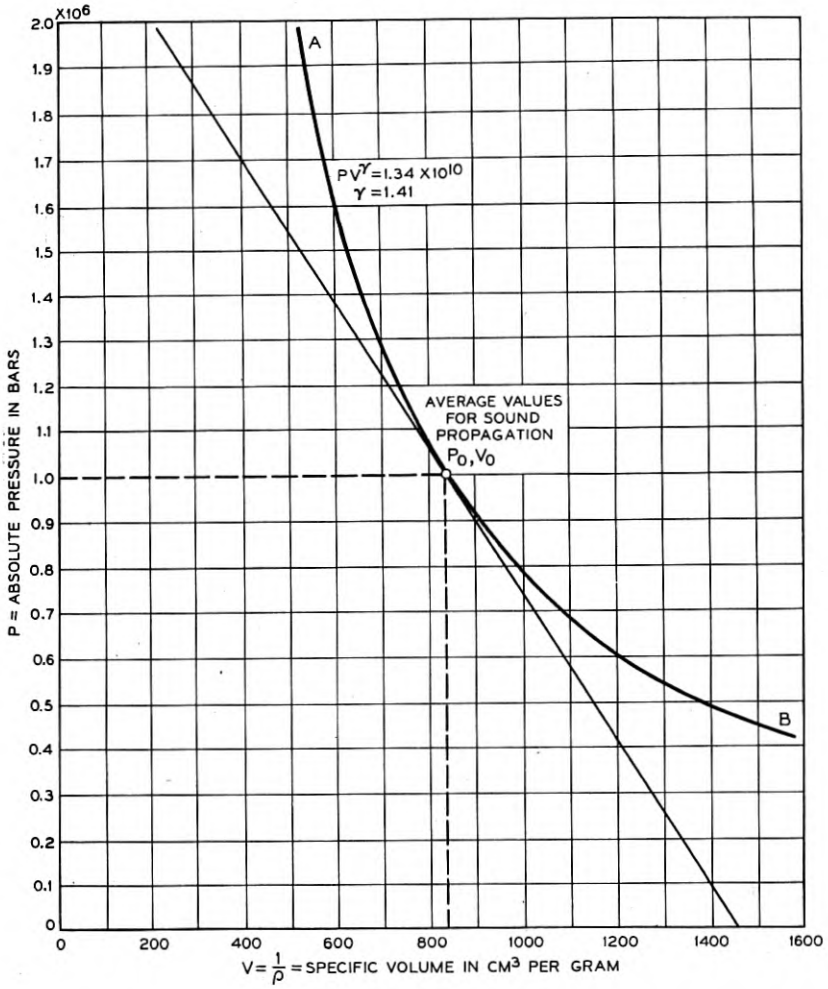


Fig. 1a—Adiabatic curve for air.

on the air; these magnitudes also increase with distance from the source and with the product of the fundamental pressures and, respectively, with the sum and difference frequencies.

THEORY OF PROPAGATION OF PLANE WAVES OF FINITE AMPLITUDE

The derivation of the exact differential equation for sound wave propagation in air involves the continuity equation, Newton's force equation and the equation expressing the relation between pressure and specific volume in a gas. Since there may be some question as to the accurate definition of the density and force in the equation of motion a somewhat detailed discussion of this subject will be given.

Following Rayleigh,[2] let $y$ and $y + (\partial y/\partial x)dx$ be the actual distances at time $t$ from the plane $x = 0$ to neighboring layers of air whose undisturbed positions are defined by $x$ and $x + dx$, respectively, Fig. 1b.

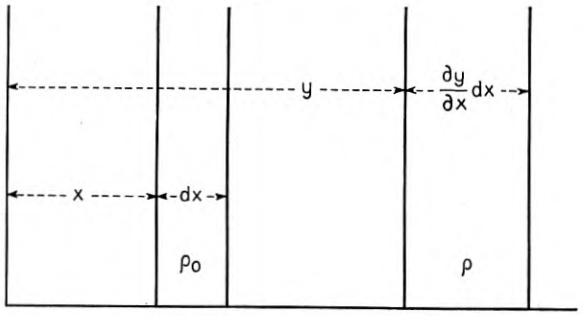


Fig. 1b.

The displacement corresponding to $y$ is thus $\xi = y - x$ and the equation of continuity of the fluid is

$$\rho = \rho_0(\partial y/\partial x)^{-1} = \rho_0(1 + \partial\xi/\partial x)^{-1}, \tag{1}$$

where $\rho$ and $\rho_0$ are the densities of the fluid in the disturbed and undisturbed states, respectively. If the effect of viscosity is neglected the exact equation of motion of the element of mass $\rho(\partial y/\partial x)\cdot dx$ is

$$\frac{\partial^2 y}{\partial t^2}\cdot\rho\frac{\partial y}{\partial x}dx = \frac{\partial^2 y}{\partial t^2}\cdot\rho_0 dx = -\frac{\partial p}{\partial y}\cdot\frac{\partial y}{\partial x}dx,$$

or

$$\rho_0(\partial^2\xi/\partial t^2) = -\partial p/\partial x, \tag{2}$$

$p$ is the pressure at the point $y$ (Fig. 1b) which moves with the air particle, not the pressure at a fixed point. Except for very large displacements these pressures are nearly the same. From equations (1) and (2)

$$\partial^2\xi/\partial t^2 = (dp/d\rho)\cdot(1 + \partial\xi/\partial x)^{-2}\cdot(\partial^2\xi/\partial x^2). \tag{3}$$

[2] Lord Rayleigh, "Theory of Sound," 2nd Ed., Vol. II, p. 31.

By virtue of equation (1), equation (3) is linear in $\xi$ only if $dp/d\rho = K\rho^{-2}$ or $dp/dv = -K$, where $v = 1/\rho$ = specific volume and $K$ is a constant. This condition is not satisfied during any ordinary variations of state of a gas, but is approximately satisfied when the variations are very small. For isothermal changes we have $pv = p_0v_0$ and for adiabatic changes:

$$p/p_0 = (v_0/v)^\gamma = (\rho/\rho_0)^\gamma, \tag{4}$$

where $\gamma$ is the ratio of the specific heats and $p_0$ is the undisturbed atmospheric pressure. In either case, for very small variations, the $pv$ curve is practically identical with the tangent to the curve, hence $dp/dv$ is practically constant (Fig. 1a).

From equations (1), (3), (4) we obtain the exact equation of adiabatic plane wave motion in a non-viscous fluid:

$$\partial^2\xi/\partial t^2 = c^2(1 + \partial\xi/\partial x)^{-\gamma-1}(\partial^2\xi/\partial x^2), \tag{5}$$

where $c^2 = \gamma p_0/\rho_0$. This equation is given by Rayleigh.[2, 3] Rocard [4] was first to call attention to the generation of harmonics in the air within an exponential horn. His theoretical solution is based on a plane wave equation in which the term $\partial^2\xi/\partial t^2$ was replaced by

$$\frac{\partial^2\xi}{\partial t^2} + \frac{\partial\xi}{\partial t}\cdot\frac{\partial}{\partial x}\left(\frac{\partial\xi}{\partial t}\right). \tag{6}$$

In support of this substitution Rocard cites Riemann's [5] treatment of the problem. However, Riemann's analysis is based on the Eulerian form of the hydrodynamical equations whereas equation (5) is derived from the Lagrangian equations. (For a comparison of these systems of equations see Lamb.[6]) In the Lagrangian notation $\partial\xi/\partial t$ and $\partial^2\xi/\partial t^2$ are the *exact* values of the velocity and acceleration, respectively, of the particle whose displacement from its equilibrium position ($x$) is $\xi$. It is to be noted that in equation (2) the term $\rho_0$, or undisturbed density, does not represent an approximation.

A rigorous solution of (5) for the displacement $\xi$ as an explicit function of $x$ and $t$ has not been obtained. As a first approximation to equation (5) we take

$$\frac{\partial^2\xi}{\partial t^2} = c^2\frac{\partial^2\xi}{\partial x^2} - (\gamma + 1)c^2\frac{\partial\xi}{\partial x}\cdot\frac{\partial^2\xi}{\partial x^2}. \tag{7}$$

[3] Lamb, "Dynamical Theory of Sound," 2nd Ed., p. 182.

[4] Y. Rocard, "Sur la propagation des ondes sonores d'amplitude finie," *Comptes rendus*, **196**, 161 (1933).

[5] Riemann, "Ueber die Fortpflanzung ebener Luftwellen von endlicher Schwingungsweitte," *Göttingen Abhandlungen*, No. 8, 1860.

[6] Lamb, "Hydrodynamics," 6th Ed., Chapter I.

This approximation restricts the dilatation $\partial\xi/\partial x$ to values small compared with unity or the excess pressure to values small compared with $\gamma p_0$, but the restriction need not be as severe as would be required for the linear approximation:

$$\partial^2\xi/\partial t^2 = c^2\partial^2\xi/\partial x^2.$$

By a method of successive approximations, carried to the second approximation, Lamb [3] derives the solution of (7):

$$\xi = a\cos\omega(t - x/c) + \frac{\gamma+1}{8}\frac{\omega^2}{c^2}a^2x[1 - \cos 2\omega(t - x/c)] \quad (8)$$

corresponding to a motion $\xi = a\cos\omega t$ imposed on the air at $x = 0$, and assuming complete absence of reflection. By virtue of (4) and (1) we have for the pressure:

$$p = p_0(1 - \gamma\partial\xi/\partial x + \cdots).$$

Neglecting the terms of small amplitude, in the region where $4\pi x$ is large compared with the wave-length $\lambda$ we have

$$p = P_{dc} + p_1 + p_2, \quad (9)$$

where

$$P_{dc} = p_0 - \gamma p_0\cdot((\gamma+1)/8)\cdot(\omega^2/c^2)a^2,$$
$$p_1 = -\gamma p_0\cdot(\omega a/c)\sin\omega(t - x/c),$$
$$p_2 = \gamma p_0\cdot((\gamma+1)/4)\cdot(\omega^3/c^3)\cdot a^2x\sin 2\omega(t - x/c),$$

or

$$p_1 = 2^{\frac{1}{2}}P_1\cos[\omega(t - x/c) + \pi/2],$$
$$p_2 = 2^{\frac{1}{2}}P_2\cos[2\omega(t - x/c) - \pi/2], \quad (10)$$

where

$$P_1 = \gamma p_0\omega a/2^{\frac{1}{2}}c, \quad (11)$$

$$P_2 = \frac{\gamma+1}{2(2)^{\frac{1}{2}}}\cdot\frac{P_1^2}{\gamma p_0}\cdot\frac{\omega x}{c}. \quad (12)$$

$P_1$ and $P_2$ are thus the r.m.s. fundamental and second harmonic pressures, respectively.

Lamb [3] also gives a solution of (7) for the case when the forced motion at $x = 0$ is: $\xi = \xi_A\cos\omega_A t + \xi_B\cos\omega_B t$. In addition to the two fundamentals and two second harmonics, the pressure now includes components whose frequencies are, respectively, the sum and difference of the two primary frequencies:

$$p_s = 2^{\frac{1}{2}}P_s\cos[(\omega_A + \omega_B)(t - x/c) - \pi/2],$$
$$p_d = 2^{\frac{1}{2}}P_d\cos[(\omega_A - \omega_B)(t - x/c) + \pi/2],$$

where

$$P_s = \frac{\gamma + 1}{2(2)^{\frac{1}{2}}} \cdot \frac{P_A P_B}{\gamma p_0} \cdot \frac{(\omega_A + \omega_B)x}{c}, \tag{13}$$

$$P_d = \frac{\gamma + 1}{2(2)^{\frac{1}{2}}} \cdot \frac{P_A P_B}{\gamma p_0} \cdot \frac{(\omega_A - \omega_B)x}{c}, \tag{14}$$

and $P_A$, $P_B$ are the two r.m.s. fundamental pressures.

If we extend Lamb's method of solution of equation (7) to the third approximation and again consider the case $\xi = a \cos \omega t$ at $x = 0$ we find that in the region $kx \gg 1$, the r.m.s. third harmonic pressure is

$$P_3 = \frac{3}{16}\left(\frac{\gamma + 1}{\gamma p_0} \cdot \frac{\omega x}{c}\right)^2 P_1^3. \tag{15}$$

In the case of the greatest r.m.s. fundamental pressure used in the experiments, $P_1 = 8000$ bars at 600 c.p.s., equation (15) indicates that the third harmonic at 400 cm from the source is about 10 db below the second harmonic.

An approximate correction for the effect of attenuation in a tube caused by viscosity and heat conduction can be obtained by assuming that each of the extraneous frequencies and the fundamental is attenuated as if it were the only wave present. Thus the r.m.s. value of the fundamental at any point $x$ is assumed to be

$$P_1 = P_0 e^{-\alpha_1 x} \qquad \text{or} \qquad dP_1/dx = -\alpha_1 P_1,$$

where $P_0$ is the r.m.s. value of the fundamental at the point $x = 0$ and $\alpha_1$ is the measured attenuation factor for the fundamental. If $P_2$ is the r.m.s. value of the second harmonic at the point $x$ and $\alpha_2$ is the measured attenuation factor for the second harmonic in the absence of the fundamental, we have by using equation (12):

$$dP_2/dx = KP_1^2 - \alpha_2 P_2 = KP_0^2 e^{-2\alpha_1 x} - \alpha_2 P_2, \tag{16}$$

where

$$K = \frac{\gamma + 1}{2(2)^{\frac{1}{2}}} \cdot \frac{1}{\gamma p_0} \cdot \frac{\omega}{c}.$$

When $\alpha_1 = 0$ and $\alpha_2 = 0$, equation (16) is equivalent to (12). The solution of (16) which is consistent with the fact that the second harmonic vanishes at $x = 0$ is

$$P_2 = [KP_0^2/(2\alpha_1 - \alpha_2)][e^{-\alpha_2 x} - e^{-2\alpha_1 x}].$$

Hence

$$\frac{P_2}{P_1} = \frac{\gamma + 1}{2(2)^{\frac{1}{2}}} \cdot \frac{P_0}{\gamma p_0} \cdot \frac{\omega}{c} x \cdot R, \tag{17}$$

where

$$R = \frac{e^{-(\alpha_2-\alpha_1)x} - e^{-2\alpha_1 x}}{(2\alpha_1 - \alpha_2)x} = 1 - \frac{\alpha_2 x}{2} + \cdots.$$

For the tube used in these experiments $\alpha_2$ may be taken as $2\frac{1}{2}\alpha_1$ which gives

$$R = 1 - \alpha_1 x/2\tfrac{1}{2} + \cdots.$$

Similar correction factors were derived for the other extraneous frequencies measured in the experiments.

## Measurements of Plane Waves of Finite Amplitude in a Tube

The experimental work consisted of measuring the second harmonic generated along a tube. Measurements were also made of the sum and difference tones when two fundamental frequencies of equal pressure were simultaneously impressed on the air of the tube. For the fundamental pressures and distances used in the experiments, the magnitudes of the other harmonics and higher order sum and difference frequencies were probably small. For high fundamental frequencies and pressures, however, these other tones are important, since they increase more rapidly with frequency and pressure than the second harmonic; for instance, the third harmonic pressure increases as the square of the fundamental frequency and as the third power of the fundamental pressure.

A sinusoidal displacement, uniform over the cross section, was impressed on the air at one end of a long tube. The tube had an inside diameter of 3.8 cm and was 1566 cm long. Measurements were made in the first 705 cm only and the remainder of the tube was used for obtaining a non-reflective termination.

A search transmitter, comprising a small tube of 0.08 cm inside diameter and 7.5 cm long, coupled to a small condenser microphone,[7] was used for the measurements. Attenuation in this search tube was sufficiently high to prevent either overloading the microphone or altering the sound wave propagated within the long tube. The search transmitter was connected to a stage of amplification so operated as to preclude non-linear distortion. This was followed by a band-pass filter which selected the frequency desired in the measurements. The filter was terminated by a measuring circuit consisting of a high-gain amplifier and a vacuum tube voltmeter. A diagram of the arrangement is shown in Fig. 2.

[7] H. C. Harrison and P. B. Flanders, "An Efficient Miniature Condenser Microphone System," *Bell Sys. Tech. Jour.*, **11**, 451 (1932).

To obtain reliable measurements throughout the length of the test tube it was necessary to reduce standing waves to a negligible magnitude. This was accomplished by laying a strip of felt in the last 761 cm of the tube, terminating the end with an acoustic resistance approximately equal to the characteristic impedance of the tube, and carefully sealing up all joints along the tube. The pressure variation in the standing wave was $\pm$ 0.3 db, which corresponds to a reflection coefficient of 0.035.

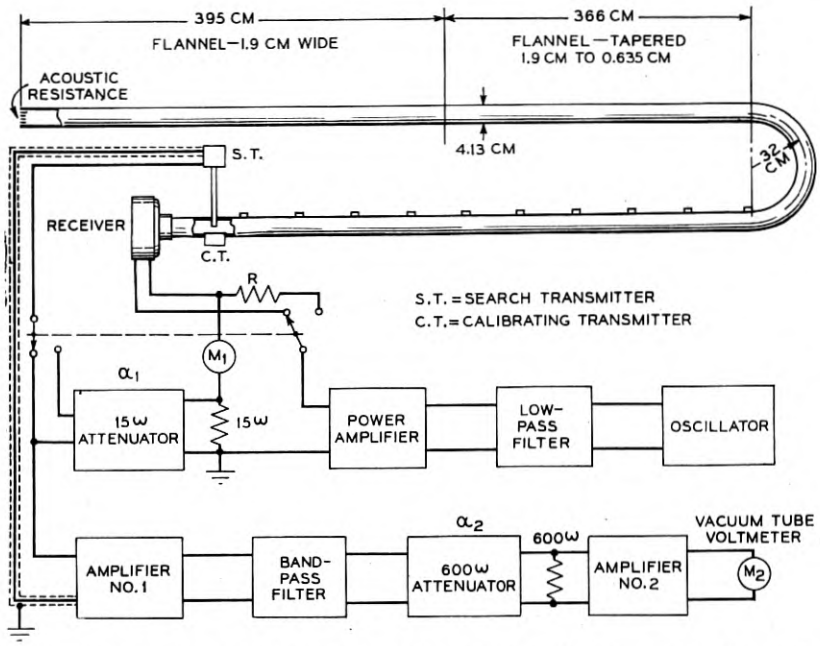


Fig. 2—Apparatus for measuring extraneous frequencies generated in air carrying intense sound waves. *R*, resistance substitute for receiver.

The oscillator current was supplied to the loud speaker through a low-pass filter and the measured harmonic content was found to be 73 db below the fundamental. Pressure measurements in the tube close to the loud speaker indicated that the harmonics generated in the measuring circuit and loud speaker were more than 50 db below the fundamental pressure at 2000 bars.

A calibration of the search transmitter was obtained by comparison with a small condenser microphone whose diaphragm was exposed directly to the sound wave at the same position on the test tube, see Fig. 2. The calibrating microphone (*C.T.*, Fig. 2) had been previously

calibrated by a thermophone.[8] The measuring circuit following the search transmitter was calibrated for each frequency measured by introducing the oscillator current into the search transmitter circuit through the attenuator ($\alpha_1$, Fig. 2).

The ratio of the pressure of the frequency generated along the tube to the fundamental pressure was measured by the attenuator $\alpha_2$, Fig. 2, at various holes along the tube in which the search transmitter was inserted. If an appreciable fraction of the harmonic generated along the tube is reflected at the end of the tube, the magnitude of the reflected component near the source may be comparable with or larger than the harmonic generated between the source and the point in question. This was found to be the case when several measurements were made near the source over a distance covering a wave-length. The measured variation in the total pressure of the harmonic over this distance was $\pm$ 2.5 db whereas the variation for a fundamental of this frequency, as previously stated, was $\pm$ 0.3 db. Therefore the measurements close to the receiver may be inaccurate.

Figure 3 shows the measured pressure ratio of the generated second harmonic to the fundamental along the tube and the theoretical curve

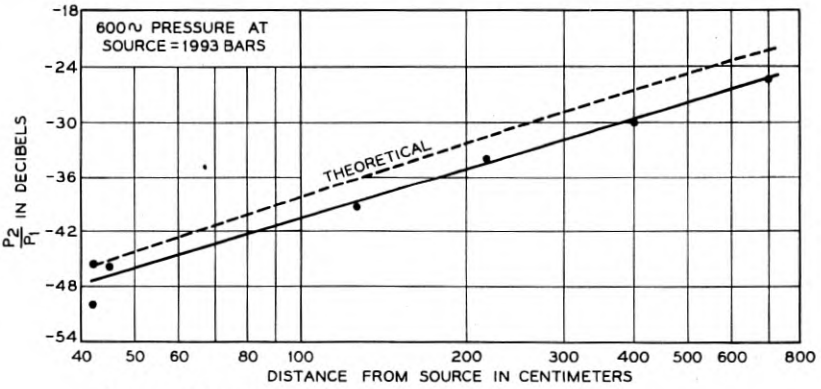


Fig. 3—Magnitude of 2nd harmonic *vs.* distance from source.

calculated from equation (17). Each of the three experimental points plotted at about 45 cm from the receiver is the average pressure ratio for a series of readings taken over a distance of a wave-length along the tube.

The measured and theoretical pressure ratios of the second harmonic to the fundamental are shown as a function of frequency and pressure in Figs. 4 and 5.

[8] L. J. Sivian, "Absolute Calibration of Condenser Transmitter," *Bell Sys. Tech. Jour.*, **10**, Jan. (1931).

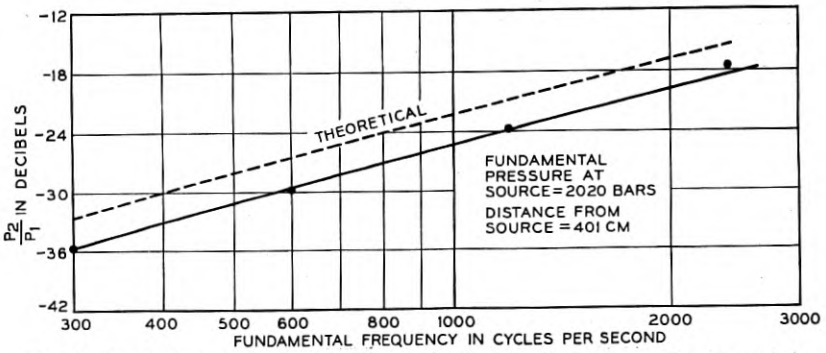


Fig. 4—Variation of 2nd harmonic magnitude with frequency of fundamental.

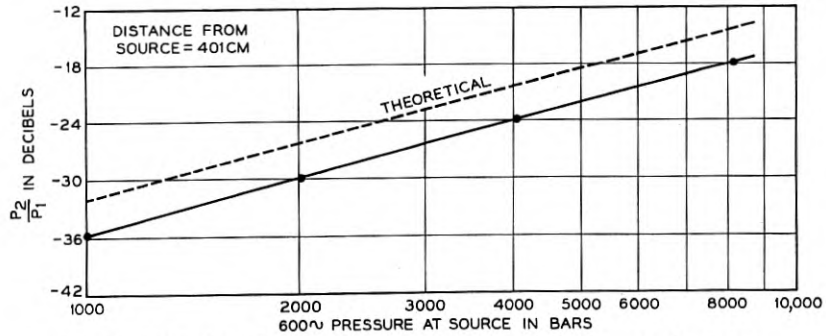


Fig. 5—Variation of 2nd harmonic pressure with fundamental pressure.

Figure 6 shows the magnitude of the first order sum and difference frequencies along the tube when two frequencies of equal pressure are simultaneously impressed on the air in the tube.

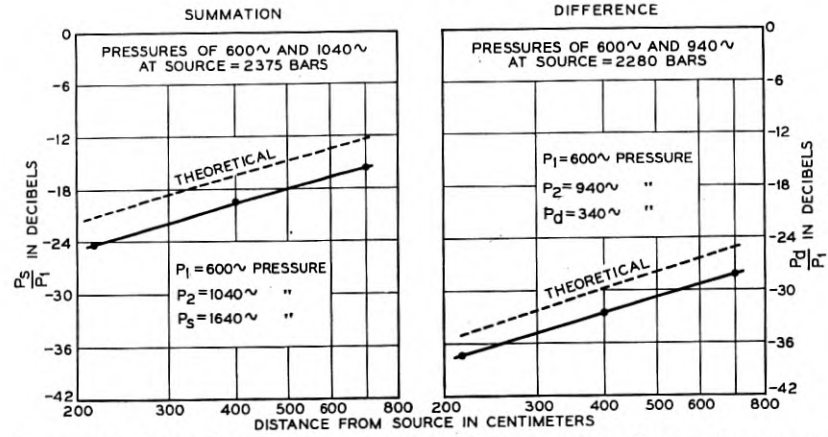


Fig. 6—Magnitude of summation and difference frequencies *vs.* distance from source.

## Exponential Horn Theory

The second harmonic generated in any short section of a horn is approximately the same as that generated in a tube of area equal to the mean area of the section of the horn. Therefore, from the tube equation and the expression for the change in pressure due to the divergence of the horn the magnitude of the generated second harmonic pressure at any point along the horn can be obtained.

The r.m.s. value of a small excess pressure in an exponential horn of section $S = S_0 e^{mx}$ is attenuated according to the law

$$P = P_t e^{-mx/2} \qquad \text{or} \qquad dP/dx = -mP/2,$$

where $P_t$ is the r.m.s. pressure in the throat of the horn (at $x = 0$). The index of taper $m$ is equal to $4\pi f_c/c$ where $f_c$ is the cut-off frequency of the horn.

From equation (12), the rate at which the r.m.s. value of the second harmonic increases along a tube is

$$\frac{dP_2}{dx} = \frac{\gamma + 1}{2(2)^{\frac{1}{2}}} \frac{P_1^2}{\gamma p_0} \cdot \frac{\omega}{c} = KP_1^2.$$

If it is assumed that the same expression represents the rate of *generation* of second harmonic along a horn, and that both the fundamental and second harmonic diverge in the same manner, the complete differential equation for $P_2$ becomes

$$\frac{dP_2}{dx} = KP_1^2 - \frac{m}{2} P_2 = KP_{1t}^2 e^{-mx} - \frac{m}{2} P_2,$$

where $P_1$ and $P_{1t}$ are, respectively, the r.m.s. fundamental pressures at the point in question ($x$) and in the throat of the horn. The solution consistent with the condition $P_2 = 0$ at $x = 0$ is

$$P_2 = [KP_{1t}^2/(m/2)][e^{-mx/2} - e^{-mx}]. \tag{18}$$

Since $P_1 = P_{1t} e^{-mx/2}$, the ratio of the second harmonic pressure to the fundamental pressure at any point $x$ in the horn is thus

$$\frac{P_2}{P_1} = KP_{1t} \cdot \frac{1 - e^{-mx/2}}{m/2} = \frac{\gamma + 1}{2\sqrt{2}} \cdot \frac{P_{1t}}{\gamma p_0} \cdot \frac{\omega}{c} x', \tag{19}$$ *

* According to this equation the magnitude of the second harmonic pressure generated in the air of a horn is 6 db lower than that given by Rocard's equation previously published in the paper "Loud Speakers and Microphones" (Reference 1), page 264, equation 1.

where

$$x' = (1 - e^{-mx/2})/(m/2). \tag{20}$$

Equations (19) and (20) indicate that the second harmonic at the mouth of an exponential horn of length $x$ is equivalent to the second harmonic at the end of a tube of length $x'$ and of uniform section, equal to the area of the throat of the horn. Thus the second harmonic in the mouth of a horn having an index of taper $m = 0.075$ cm$^{-1}$, cut-off frequency 200 c.p.s. and length 78 cm is equal to the second harmonic at the end of a straight tube 25 cm long and of the same diameter as the throat of the horn.

## Exponential Horn Measurements

Measurements of the output of a horn attached to a moving coil receiver were made in an acoustically damped room. The horn had a throat diameter of 3.8 cm, a length of 78 cm and a cut-off frequency of 200 c.p.s. The diaphragm of the loud speaker was coupled to the throat of the horn through a straight tube 13 cm long. A filtered single frequency tone was impressed on the loud speaker and the sound was picked up by a small microphone in front of the horn. The fundamental and second harmonic voltages from the microphone amplifier were separated by means of a band-pass filter and measured.

The approximate acoustic power at the throat of the horn was calculated from the known efficiency of the loud speaker and the electrical voltage and current supplied.

The measured and calculated ratios of the second harmonic pressure to the fundamental pressure at the mouth of the horn, including the effects of generation of second harmonic in both the horn and the straight tube coupling the receiver and the throat of the horn, are shown in Fig. 7 in terms of the sound output in watts. See equations

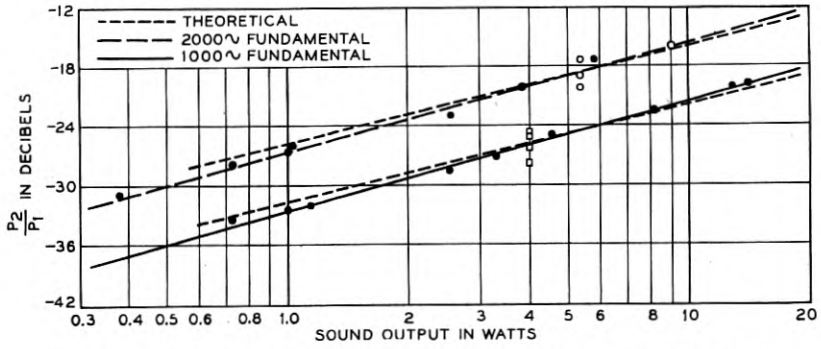


Fig. 7—2nd harmonic generated in an exponential horn *vs.* sound output.

(17) and (19). The attenuation due to viscosity and heat loss in the horn has been neglected.

A number of measurements at various microphone positions in front of the horn, shown by the dotted circles, indicate the difficulty of obtaining accurate results in a room. The average of the measurements at a number of random positions in front of the horn at a constant sound power output and the measurements at a single position for various sound outputs gives the plotted curve which is probably not greatly different from that which would be obtained in open air.

Figure 8 shows the measured and calculated ratios of the second harmonic pressure to the fundamental pressure for various fundamental frequencies.

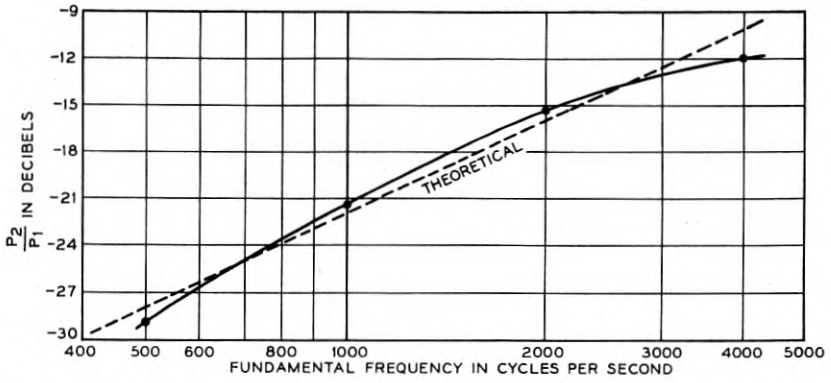


Fig. 8—2nd harmonic generated in an exponential horn *vs.* fundamental frequency (sound output = 10 watts).

## Demonstration of Extraneous Frequencies

Almost two hundred years ago Sorge, a German organist, and Tartini, an Italian violinist, discovered independently, apart from all theory, that the union of two loud independent tones produced a difference tone. That this is not entirely a subjective tone produced by the ear but is actually present in the air was demonstrated by others some years later by the use of a tuned resonator. With modern apparatus consisting of power amplifiers, oscillators and tuned electromechanical vibrators it is a relatively simple matter to show not only the difference tone but the summation and harmonic tones as well.

An exponential horn was attached to the open end of the 1566 cm tube previously described but with the damping material removed. A moving coil microphone placed in front of the horn picked up the complex tone produced when two equal pure tones of 600 and 940

c.p.s. were impressed on the air. The microphone voltage was amplified and impressed on six torsional vibrators tuned to 340, 600, 940, 1200, 1540 and 1880 c.p.s. Spherical mirrors attached to the vibrators produced on a screen bands of light the amplitudes of which were approximately proportional to the relative pressures of the various frequencies in the complex tone.

For the higher power inputs to the loud speaker used in the experiment, the presence of the sum and difference frequencies and the harmonic frequencies was easily observed. At these power outputs the quality of the sound was very disagreeable and the fundamental tones could hardly be distinguished.

## Conclusion

The theoretical and experimental determinations of the extraneous frequency waves generated in the air within a tube are in good agreement as regards the variation in magnitude with frequency, distance from the source and magnitude of the primary tones. The magnitude of the second harmonic is very nearly proportional to the distance from the source, to the fundamental frequency and to the square of the amplitude of the fundamental pressure. When there are two primary tones, the extraneous frequencies generated in the air include, as well as the harmonics of the primary tones, frequencies which are, respectively, the sum and the difference of the primary frequencies and also other higher order tones. The magnitudes of the summation and difference tones are very nearly proportional to the distance from the source, to the product of the magnitudes of the two primary pressures, and in each case, to the frequency of the particular combination tone. As regards the absolute magnitudes of the generated tones in the tube all of the measured values are about 3 db lower than the theoretical values.

Good agreement was obtained also between the experimental and theoretical determinations of the second harmonic generated in the air within an exponential horn, as regards proportionality to the fundamental frequency and power and also as to absolute magnitude. In fact the agreement in absolute magnitude was closer for the horn than for the tube, but not much significance should be attached to this fact, as the horn theory is developed from the theoretical solution for the tube and the horn measurements are known to be less reliable than the tube measurements.

## Abstracts of Articles from Bell System Sources

*North Atlantic Ship-Shore Radio Telephone Transmission During 1932–1933.*[1] CLIFFORD N. ANDERSON. This paper extends the analysis of ship-shore radio transmission data for an additional two-year period beyond that reported on in a previous paper. Contour diagrams show the variation of signal field with time of day and distance for the winter, summer, spring, and fall seasons and for the approximate frequencies 4, 8, and 13 megacycles.

A comparison is made with the data obtained during 1930 and 1931. In general, transmission during 1932–1933 tends to be somewhat better on frequencies below about 9 megacycles and somewhat poorer on frequencies above 9 megacycles. At 4 megacycles the increase is of the order of 10 decibels, and for 13 and 17 megacycles, the decreases are about 6 and 10 decibels, respectively.

*Loudness, Pitch and the Timbre of Musical Tones and Their Relation to the Intensity, the Frequency and the Overtone Structure.*[2] HARVEY FLETCHER. It is generally supposed that the three psychological characteristics of a musical tone, namely, loudness, pitch and timbre, are each directly dependent upon the corresponding physical characteristics of the sound wave producing the tone, namely, intensity, frequency of vibration and overtone structure. In this paper it is shown that each of the psychological characteristics is not dependent upon a single one but upon all three of the above mentioned physical characteristics. Quantitative measurements of these relationships have been made using a large number of observers. For example, it was found that one tone may have the same pitch and loudness as another and yet have an intensity 100 times as great, the difference being due to the difference in the overtone structure; or two tones may have the same pitch and overtone structure and yet have frequencies of vibration that differ as much as 5 per cent, the difference being due to the different intensity of the two tones. An empirical formula showing the dependence of loudness upon the three physical characteristics mentioned has been formulated although no such formula has been found for showing a similar relationship for pitch. Such a relationship is shown graphically for pure tones. Also some very marked effects upon

[1] *Proc. I. R. E.*, October, 1934.
[2] *Jour. Acous. Soc. Am.*, October, 1934.

the pitch of a musical tone due to the changes of overtone structure are described.

*Overvoltages on Transmission Lines.*[3] C. L. GILKESON and P. A. JEANNE. Observations of line-to-ground voltages have been made under routine operating conditions on an isolated neutral system, a Petersen coil system, 3 neutral resistance grounded systems, and 2 directly grounded systems. Results of these observations are given in this paper. Measurements were made with oscillographs supplemented, on all but 2 systems, by surge recorders.

*An Acoustic Spectrometer.*[4] C. N. HICKMAN. A series of tuned reeds are mounted so that they may be electromagnetically driven. Each reed carries a small concave mirror with which light from an illuminated slit is brought to a focus on a screen. These slit images are lined up in the order of the reed frequencies. When a current having a complex wave, such as the speech current from a microphone, is passed through the electromagnet, the reeds and in consequence the slit images on the screen will oscillate. The driving system and the reeds are so designed that the amplitude of oscillation of each image is proportional to the strength of the corresponding harmonic component in the driving current. Therefore, by observing or photographing the slit image amplitudes, the frequencies and the relative energy content of the components of a complex current may be determined. A spectrometer of this type covering a small frequency range (50 to 3109 cycles) was built for demonstration purposes. The range of such an instrument can be extended to higher and lower frequencies.

*The Measurement of Harmonic Power Output of a Radio Transmitter.*[5] P. M. HONNELL and E. B. FERRELL. A method of determining the harmonic power output of a high-frequency radio transmitter is described. It is a method for measuring the power delivered by the transmitter to the antenna system, as distinguished from the more common method of measuring harmonic field strengths at specified locations. It is essentially a comparison method. The unknown harmonic power, present with the fundamental, is compared by means of a sufficiently selective receiving set with a known comparison power which is supplied in the absence of the fundamental. The method in practice seems to be accurate within about one decibel. It is applicable to the measurement of power other than harmonic power.

[3] *Elec. Engg.*, September, 1934.
[4] *Jour. Acous. Soc. Am.*, October, 1934.
[5] *Proc. I. R. E.*, October, 1934.

*Piezoelectric Frequency Control.*[6] F. R. LACK. This paper discusses the use made of the piezoelectric effect in designing sub-standard time-keepers and frequency generators. The nature of the piezoelectric effect is outlined and mention is made of the various classes of crystals in which it is found. The technic of setting up, electrically, various types of mechanical vibrations in piezoelectric crystals is described together with methods of obtaining very high frequencies. Other applications of the piezoelectric effect, such as to loud speakers, submarine signaling, etc., are briefly reviewed.

*Reverberation Measurements in Auditoriums.*[7] G. T. STANTON, F. C. SCHMID and W. J. BROWN, JR. This paper presents some of the problems encountered in attempts to measure the reverberation time in auditoriums, and indicates certain procedures found helpful in their solution. In addition, comparison is offered between results obtained with two step-by-step methods and a continuous decay.

[6] *Jour. S. M. P. E.*, October, 1934.
[7] *Jour. Acous. Soc. Am.*, October, 1934.

## Contributors to this Issue

ALFRED C. BECK, E.E., Rensselaer Polytechnic Institute, 1927. Test Department, New York Edison Company, summers 1926–27. Instructor in mathematics, Rensselaer Polytechnic Institute, 1927–28. Member, Technical Staff, Bell Telephone Laboratories, 1928–.

EDMOND BRUCE, B.S., Massachusetts Institute of Technology, 1924. Radio service, U. S. Navy, 1917–19. Western Electric Company, 1924–25; Bell Telephone Laboratories, 1925–. Mr. Bruce has been engaged in the development of short-wave radio receivers and field-strength measuring equipment, and has specialized in directive antenna systems for short-wave radio communication. He was awarded the Morris Liebmann Memorial Prize by the Institute of Radio Engineers in 1932.

A. B. CLARK, B.E.E., University of Michigan, 1911. American Telephone and Telegraph Company, 1911–34; Toll Transmission Development Engineer, 1928–34. Bell Telephone Laboratories, 1934–. As Toll Transmission Development Director, Mr. Clark is engaged in work largely concerned with toll telephone and telegraph systems.

W. C. ELLIS, Ch.E., Rensselaer Polytechnic Institute, 1924; Ph.D., 1927. Bell Telephone Laboratories, 1927–. Dr. Ellis has been engaged in metallurgical studies on magnetic materials and copper alloys.

R. T. JENKINS, B.S., in Electrical Engineering, Cooper Union, 1917; E.E., 1920. Engineering Department, Western Electric Company, 1916–25; Research Department, Bell Telephone Laboratories, 1925–. Mr. Jenkins' work has been largely in connection with transmission testing methods and apparatus, and with electro-acoustic calibrations.

J. B. JOHNSON, B.S.. University of North Dakota, 1913; M.S., University of North Dakota, 1914; Ph.D., Yale, 1917. Engineering Department, Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. As Research Physicist, Dr. Johnson has been engaged in the development and application of special vacuum tubes and gaseous discharge devices.

M. J. KELLY, B.S., Missouri University, 1914; M.S., University of Kentucky, 1915; Ph.D., University of Chicago, 1919; Instructor of Physics, University of Kentucky, 1914–15; Research Assistant, University of Chicago, 1915–18. Western Electric Company, 1918–25; Bell Telephone Laboratories, 1925–. As Vacuum Tube Development Director, Dr. Kelly has worked on the development of thermionic and photoelectric devices.

FREDERICK B. LLEWELLYN, M.E., Stevens Institute of Technology, 1922; Ph.D., Columbia University, 1928. Western Electric Company, 1923–25; Bell Telephone Laboratories, 1925–. Dr. Llewellyn has been engaged in the investigation of special problems connected with radio and vacuum tubes.

LEWIS R. LOWRY, B.S. in Electrical Engineering, University of Washington, 1927. Pacific Telephone and Telegraph Company, summers 1922–26. Member, Technical Staff, Bell Telephone Laboratories, 1927–.

H. T. O'NEIL, B.S. in Electrical Engineering, State College of Washington, 1923. Engineering Department, Western Electric Company, 1923–25. Bell Telephone Laboratories, 1925–. Mr. O'Neil has been engaged in theoretical acoustic research.

H. A. PIDGEON, B.S., Ohio University, 1911; M.S., 1912; Ph.D., Cornell University, 1918. Western Electric Company, Engineering Department, 1918–25. Bell Telephone Laboratories, 1925–. As Vacuum Tube Development Engineer, Dr. Pidgeon is engaged in research and development work on vacuum tubes.

A. L. SAMUEL, A.B., College of Emporia (Kansas), 1923; S.B. and S.M. in Electrical Engineering, Massachusetts Institute of Technology, 1926. Instructor in Electrical Engineering, Massachusetts Institute of Technology, 1926–28. Bell Telephone Laboratories, 1928–. Mr. Samuel has been engaged in research and development work on vacuum tubes.

E. E. SCHUMACHER, B.S., University of Michigan; Research Assistant in Chemistry, 1916–18. Engineering Department, Western Electric Company, 1918–25; Bell Telephone Laboratories, 1925–. As Assistant Research Metallurgist, Mr. Schumacher is in charge of a group whose work relates largely to research studies on metals and alloys.

A. L. THURAS, B.S., University of Minnesota, 1912; E.E., 1913. Laboratory assistant with U. S. Bureau of Standards, 1913–16. Graduate student in physics, Harvard, 1916–17. Engineering Department, Western Electric Company, 1920–25. Bell Telephone Laboratories, 1925–. Mr. Thuras has been engaged in studies and development of electro-acoustic devices and instruments.