ELECTRICAL COMMUNICATION

Technical Journal of the
International Telephone and Telegraph Corporation
and Associate Companies

1944

VOL. 22 No. 1

ELECTRICAL COMMUNICATION

Technical Journal of the
INTERNATIONAL TELEPHONE AND TELEGRAPH CORPORATION
and Associate Companies

H. T. KOHLHAAS, Editor

Published Quarterly by the

INTERNATIONAL TELEPHONE AND TELEGRAPH CORPORATION

67 BROAD STREET, NEW YORK 4, N.Y., U.S.A.

Sosthenes Behn, President Charles D. Hilles, Jr., Vice President and Secretary

Subscription, $1.50 per year; single copies, 50 cents

Volume 22 1944 Number 1

CONTENTS

NEWLY COMPLETED FEDERAL TELEPHONE AND RADIO LABORATORIES' HANGAR, INCLUDING LABORATORY AND SHOP FACILITIES, LOCATED AT WESTCHESTER AIRPORT, RYE LAKE, NEW YORK.

Refined Quality Control Speeds Up Production*

Control Chart Helps Spot Process Trouble, Saves Inspection Cost

By JOHN GAILLARD

Mechanical Engineer, American Standards Association

Editor's Note: Refined quality control is gaining wide recognition as a valuable aid in weeding out assignable causes of product variation, in early detection of potential causes of trouble in the production process, and in setting limits for quality variation. Further, when quality is controlled, data obtained from samples of the product can give evidence of the greatest possible reliability in judging quality. The present article, reprinted by permission, concisely and ably summarizes the fundamentals of this timely subject.

As stated by the author, the U. S. War Department is using the control chart method of quality control extensively in the arsenals and encourages its use by private industry. This technique thus is becoming an important factor in the war effort of the U.S.A. and, also, in Great Britain, Australia and Canada, where the American War Standards in this field have been adopted by the respective national standardizing bodies.

MANY manufacturing plants engaged in war production complain that their work results in an excessive proportion of non-acceptable product or, in other words, too large a "percent defective." This may be due to any one of a number of causes. It may be that the work is different from anything the plant has done before, or that the manufacturing tolerances specified by the government are closer than those to which the plant is accustomed.

So long as trouble of this kind persists, the manufacturer will say that he has not yet got "control" of the quality of his product.

Such a condition calls for action. Although some of the defective product may be salvageable, there is still loss of material, labor, machine capacity, and—a highly important factor just now—loss of *time*.

What Is Control?

Ideally, control of quality of a manufactured product would mean that every unit of that product meets its specified requirements. Thus, perfect dimensional control of machine parts would be achieved if the sizes of all parts manufactured fell within their gage limits. In practice, no such absolute perfection exists. The percent defective may be cut down to a very low figure, but we can never reasonably expect to get no rejections whatsoever. So far, when the manufacturer has said that he had attained control of quality, he usually has meant that he had succeeded in keeping the percent defective down to a level satisfactory to him—a matter of personal judgment. If at first the percent defective was 4, and he had gradually brought it down to 0.5, he might hold that he had attained control, even though by further efforts he could have reduced the percent defective still more, say, to 0.03.

This vagueness of the "traditional" concept of control has several practical disadvantages, some of which are quite important. For example, if the manufacturer, in spite of efforts to reduce the percent defective to an acceptable figure, has failed to do so, he may ask himself whether he should keep on trying or give up the matter as impossible. If the kind of work is entirely new to him, he may easily take the wrong decision and either give up, although with further effort the problem could be solved, or, conversely, persist where success is out of the question. Therefore, the "traditional" concept of control does not give the manufacturer a definite answer to practical questions like these: Can the percent defective be reduced by adjusting the production process in use, or must this process undergo basic changes before improvement may reasonably be expected? What is the highest degree of accuracy attainable with a certain production set-up—that is, with given materials, tools, machines, operators, and methods of manufacturing and inspection? Is it practical, in a specific case, to shift from selective fitting to assembly of interchangeable parts, or are the closer manufacturing limits required in the latter case prohibitive?

* Adapted from an article published in *American Machinist*, December 10 and 24, 1942, and in *Industrial Standardization*, April and May, 1943.

Quality Control Standards

The manufacturer will get an answer to such questions by applying a technique described in a set of three American War Standards on Quality Control. Here, the concept "control" is given a definite meaning and the manufacturer is supplied with a criterion for detecting lack of control. Also, where such lack is indicated, he knows that he must look for trouble in the production process so that he may find its cause with the least possible delay and eliminate it, if practicable. However, if there is no indication of lack of control, the manufacturer may safely assume—as wide practical experience has shown—that he is doing the best he can, with the process in use. Therefore, if this "best" is not good enough to meet the requirements of the product—specified, for example, in terms of manufacturing limits—he knows that he will be able to get product of the right quality only by making basic changes in the process, and not by merely adjusting it.

Origin of Standards

Upon request by the War Department, the American Standards Association (ASA) started in 1940 a project on the Application of Statistical Methods to the Quality Control of Materials and Manufactured Products. Three American War Standards, completed by an ASA committee,[1] have so far been published. They are: Guide for Quality Control (Z1.1-1941); Control Chart Method of Analyzing Data (Z1.2-1941); and Control Chart Method of Controlling Quality During Production (Z1.3-1942).[2]

What Is Quality?

Before we can discuss *control* of quality, we must have a clear picture of what we understand by *quality*. Roughly, the quality of a product may be defined as its suitability for a given purpose. Hence, to be able to judge or measure quality, we must begin by specifying the requirements we want the product to meet or, in other words, the *quality characteristics* we want the product to possess. Such a characteristic may be dimensional accuracy specified in terms of the manufacturing limits within which certain dimensions of a workpiece must be held if it is to be accepted by the inspector. Other specified quality characteristics may be the tensile strength and the elongation of the steel of which the piece is made.

Specification of quality is most definite if the characteristics involved are stated in terms of measurement. If this is impossible, we may at least be able to establish a sample or model to serve as a basis of comparison for the product to be made. An example is the finishing of a workpiece to a "surface quality" represented by a sample block or disk.

Collective Quality

So far, we have discussed the concept "quality" as relating to a *unit* of product—that is, to an individual piece, subassembly, or complete article for which quality requirements have been specified. However, when we think of *control* of quality, we do not have in mind whether a given unit of product meets its requirements but, rather, whether or not the totality of the product, as it comes in lots or in a continuous flow from the line, is satisfactory in the sense that it contains a sufficient percentage of acceptable units. In other words, the product is considered here as a collection of units and the term "quality control" accordingly refers to the *collective quality* of the product. If all of the pieces in a lot of 1,000 are inspected and eight pieces are rejected, we can express the collective quality of this lot by saying that its percent defective is 0.8.

How small should be the percent defective in a specific case to justify our claiming that we have got control of quality? Is there any criterion that will aid us in answering this question?

Level and Dispersion of Quality

Before we go more deeply into this question, let us consider the concept "collective quality." Suppose we look at ten lots of 1,000 pieces each, and determine the percent defective of each lot by gaging all of its units. The collective quality

[1] Membership: H. F. Dodge, Bell Telephone Laboratories, *chairman;* A. G. Ashcroft, Alexander Smith and Sons Carpet Co.; W. Edwards Deming, Bureau of the Census; Leslie E. Simon, Army Ordnance Dept.; R. E. Wareham, General Electric Co.; and John Gaillard, ASA, *secretary*.

[2] The first two standards were published in a single pamphlet in 1941 and the third standard was published in July, 1942. Copies are available from the ASA, 29 West 39 St., New York 18, N. Y.

of these ten lots, considered together, may then be measured by the arithmetic mean or *average* of the individual percentages of the ten lots. This average represents their common *level of quality* which we would have found also by combining the ten lots into a large one of 10,000 pieces and determining the latter's percent defective. However, when we have ten separate lots of 1,000 pieces, we come upon another aspect of their collective quality. This is the extent to which the qualities of the 1,000-piece lots deviate from their common quality level (average percent defective). If their qualities all are rather close to this common level, this is an indication that the overall quality of the product has a high degree of uniformity. If, on the contrary, the qualities of the 1,000-piece lots are spread out or dispersed more widely, the overall quality appears to be less uniform. Therefore, the *dispersion* of individual qualities or group qualities about their common level of quality is a second important characteristic of collective quality. It gives a picture of the degree of uniformity of quality in the total volume of product under consideration.

Measurement of Collective Quality

To *measure* collective quality, we must have units of measurement for the *level* and the *dispersion*. We have already seen that the level can be simply expressed by the *average* of the measurements (observations) of a given quality characteristic. If a quality characteristic X, such as the size of a workpiece or the tensile strength of a specimen of material, is measured on each of a number of units n, and the observations are: $X_1, X_2, \cdots X_n$, their average $\bar{X}$ (pronounced "X bar") is expressed by the formula:

$$\bar{X}=\frac{X_1+X_2+\cdots+X_n}{n}.$$

To express the *dispersion*, we can use several kinds of measures. One of those commonly used in problems of this kind is the *standard deviation* (root-mean-square deviation) of the observations from their average. It is customary to designate the standard deviation by the Greek letter σ (sigma). In the example just mentioned, the standard deviation of the n observations from their average $\bar{X}$ is expressed by the formula:

$$\sigma=\sqrt{\frac{(X_1-\bar{X})^2+(X_2-\bar{X})^2+\cdots+(X_n-\bar{X})^2}{n}}$$

Another measure of dispersion frequently used here is the *range* (R). This is the difference between the highest value and the lowest value among a set of observations, or

$$R=X_{\max}-X_{\min}.$$

For example, we may take 25 samples of 100 pieces each, all of these being gaged, and determine the percent defective for each sample. If then the maximum percent defective observed is 5, while the minimum is 2, the range of percent defective for these 25 samples is 3.

Other statistical measures used to express characteristics of collective quality are the *fraction defective* (p), which is the ratio of the number of defective units to the total number of units under consideration;[3] the *number of defectives* (pn) in a sample of n units; and the *number of defects* (c) in a sample of stated size, such as the number of flaws in 100 feet of wire. Which statistical measure or measures should be used must be decided in each individual case.

Inspection by Sampling

When a quality characteristic is measured on each unit of a lot of product, it is a simple matter to express the collective quality of the lot in such measures as the average, the standard deviation, and the range, since these values can be computed from the observations made on the units. The problem becomes more complicated when 100 percent inspection is impracticable and the collective quality of a lot must be determined, with the closest possible approximation, from observations made on samples drawn from that lot. Inspection by sampling may be necessary because the inspection test is destructive (for example, the tensile test made on a specimen of material, or the firing test of ammunition),—or because 100 percent inspection, while technically feasible, is uneconomic (for example, where parts come in a large flow from an automatic machine). In such cases we can get a picture of lot quality only from the combined pictures of a number of sample qualities—each of which is in its turn a composite picture of a number of unit qualities.

[3] The use of *percent defective* ($100p$) is often preferred.

In sampling inspection, the important question arises:

To what extent does a sample taken from a lot of product present a reliable picture of the quality of the entire lot? The answer depends on a number of factors, such as the sizes of the lot and the sample, and the degree of uniformity of quality in the lot. Since the make-up of a sample, and hence the relation between sample quality and lot quality, depends on *chance*, valid conclusions regarding lot quality can be drawn from sample qualities only with the aid of the theory of probability. This fact and, in general, the statistical background of the standards under discussion, need not dismay the technical expert who is not conversant with statistical theory. The practical tool to be used by the man in charge of quality control, which was developed through cooperation between the engineer and the statistician and described in the standards, requires merely the use of simple arithmetic. This tool is the *quality control chart.*

State of Statistical Control

The quality control chart is based on the concept "statistical control" which will be explained briefly here.

The variations in the quality of a product, from unit to unit, or from sample to sample, or again, from lot to lot, are due to numerous causes. From the viewpoint of quality control, these causes may be divided into two classes: (1) causes of variation whose effect is insignificant and which, therefore, merit no investigation, and (2) causes of variation whose effect is significant in that they result in excessive fluctuations in quality and, hence, must be eliminated if we are to get control. Causes of variation of the latter type are called *assignable causes*,—which practically means causes of trouble. When the latter are found and eliminated from the production process, and the variation in quality has become insignificant, the production process is said to be in a state of statistical control or, briefly, *in a state of control.* The residual variation in quality is due solely to the combined effect of unknown causes which by their very nature cannot be weeded out systematically. Possibly, with increasing knowledge, some of these unknown causes may become assignable in the course of time, in which case their elimination will still further reduce the residual variation. However, for a given process, considered at a certain time, the amount of knowledge at our disposal determines the limit to which our finding of assignable causes and their elimination can go.

It will be clear that this approach to the problem of quality control becomes workable in practice only if we have a criterion to determine whether a variation in quality is *significant.*

Normal Distribution

Such a criterion is found by making a comparison between the unknown distribution of quality, in which we are interested in a specific case, and a known theoretical distribution representing a "model" of perfect statistical control—that is, a distribution resulting from a condition where no assignable causes of variation are present. Such a comparison enables us to decide whether the observed distribution appears to deviate from the "model" distribution to such an extent as to indicate lack of control. This procedure is similar to the use of an optical comparator for checking the contour of a workpiece, such as a screw or a gear, against an ideal contour. In the quality control chart technique, we use as the model of perfect control the so-called *normal distribution* which is graphically represented by the normal frequency curve, or Gauss curve, shown in diagram A, Fig. 1. Its abscissae represent here the values of a certain quality characteristic observed on the individual units of a lot. Its ordinates represent the respective frequencies with which these quality values (observations) occur in the lot. The area included between the curve and the horizontal axis represents the total number of observations. A mathematical property of this frequency curve is that approximately 99.73 percent of the total observations fall within two limits located from the central line $\bar{X}'$ at equal distances $3\sigma'$, in which σ' is the standard deviation of the observations from the central line. Such limits are called "3-sigma" limits.

The diagrams B and C, Fig. 1, show unsymmetrical distribution curves. For both of these, the percentage of observations that falls outside the 3-sigma limits is larger than in diagram A.

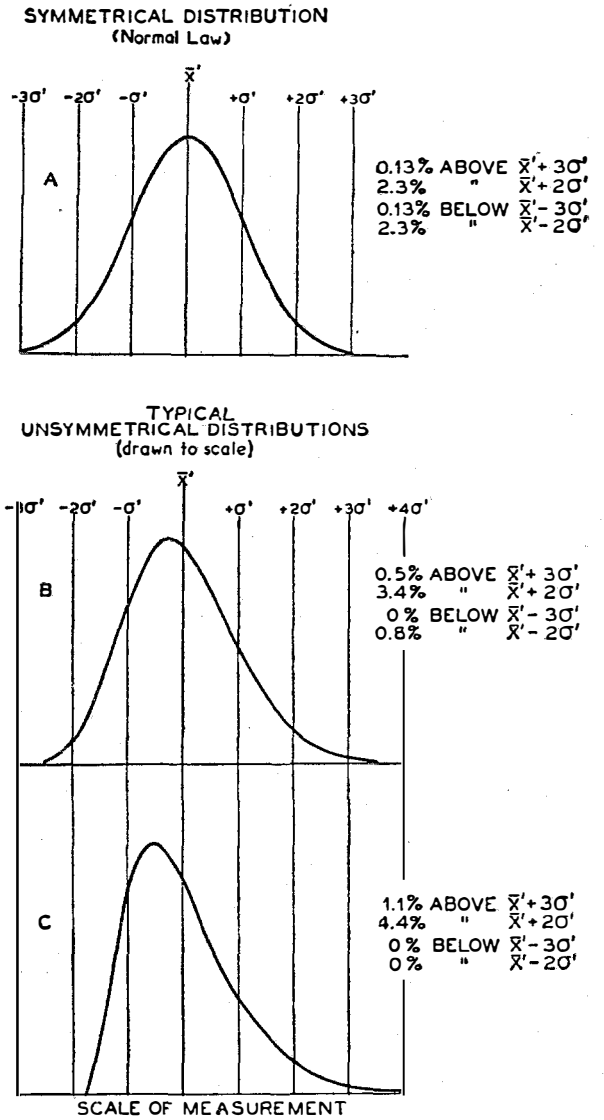

Fig. 1—Showing percentages at the far ends of distributions having different shapes.

The Quality Control Chart

The quality control chart is a graphical record of quality having two control limits placed in a manner to be explained below. Data obtained from quality observations made on samples are plotted on the chart. If a plotted point falls *outside* the control limits, this should be taken as an indication of the presence of an assignable cause of variation (cause of trouble) in the production process. Thus, the quality control chart gives us a picture of the condition in the process and warns us when we must hunt for trouble. In this way, the control limits help us to classify causes of variation into the two groups mentioned—assignable causes, and causes that merit no investigation. Such a criterion, which aids us in deciding whether or not there appears to be lack of control, is absent in what was called, in the beginning of this article, the "traditional" method of quality control.

Placing the Control Limits

Suppose we want to judge whether or not a product comprising a flow of similar units shows lack of control. As we measure the quality of individual units, we obtain a collection of measurements that form a sequence with time, and the way these measurements behave can provide us with a picture of the manufacturing process behind the scenes. Whether we are measuring each and every unit or only a fractional part of them, we can arrange the measured values in groups, in a time sequence. Each such group might represent all of the units turned out, say, in an hour's time, or again it might correspond to a sample of units for that period. In either case it constitutes a "sample of the process." If for each of these samples we compute the average of the measured values and the standard deviation of the measured values about their average, we have a succession of average values and of standard deviation values which should show a certain degree of stability if the process is controlled. What is needed is a quantitative criterion of this kind of stability—limits of expected variation of such averages and of standard deviations. By means of the formulas and tables given in the standards under discussion, we can compute in a simple manner, for any sample size n (number of units in a sample), the 3-sigma limits that are to serve as control limits on the quality control chart. If quality (the result of the process) is controlled, we shall expect to find that practically all [4] of the averages fall *within* the 3-sigma limits. If it is not controlled, however, we shall expect a number of these values to fall *outside* the 3-sigma limits.

[4] By "practically all" we mean this: Since 99.73 percent of the observations are expected to fall inside the 3-sigma limits, only 0.27 percent of the observations, or about 3 observations in 1,000, would be expected to fall outside these limits.

A similar procedure can be followed to obtain control limits for the range (R), the fraction defective (p), the number of defectives per sample (pn), or the number of defects per sample (c). Recommendations for the use of one or more of these statistical measures in a specific case are given in the standard Z1.3-1942.

It will thus be seen that the new standards, developed through cooperation between the engineer and the statistician, enable the man in charge of quality control to construct a set of control limits that help him answer the question, "Is there lack of control?" in the same way as a set of gage limits help the inspector of the product answer the question, "Is this piece acceptable?"

Control Limits and Specification Limits

Control limits are basically independent of the specification limits for a product. The latter, in the mechanical industry commonly called "manufacturing limits," are specified by the designer to indicate what is required, in his opinion, to make the product function correctly and to give it a reasonable length of wear life. Therefore, each unit of product (workpiece) must meet these limits if it is to be accepted by the inspector.

Control limits, on the contrary, are based on data obtained by inspecting units of product.

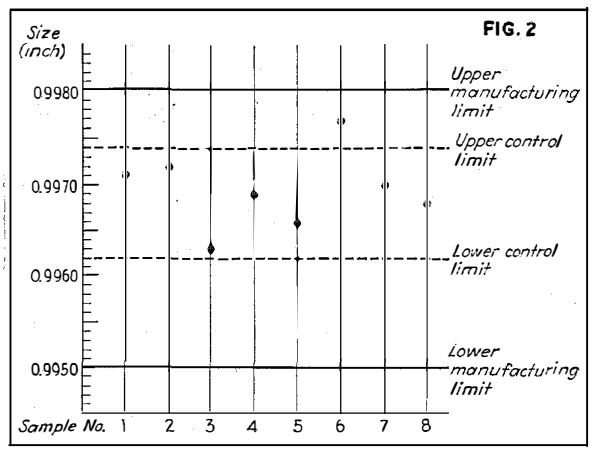

FIG. 2

Fig. 2—Chart showing "manufacturing limits" specified for the quality required of each unit of product and "control limits" against which are plotted average qualities of samples taken and inspected while the product is being manufactured.

These are statistical limits which represent, not what is desired of the product, but rather what happened while the product inspected was being made.

The difference between specification limits and control limits is illustrated by Fig. 2 which refers to a workpiece whose outside diameter must be held between 0.9950 and 0.9980 inch. After a control chart has been kept for some time by taking samples of 4 pieces, the control limits computed for a certain period are 0.9962 and 0.9974 inch. They fall within the specification limits and so long as the plotted points remain inside the control limits everything may be assumed to run smoothly. The point plotted for sample No. 6 warns of trouble in the process. Yet, this does not necessarily mean that the quality of the product has been impaired to such a degree that there are rejections in the inspection department. Point No. 6 (sample average 0.9977 inch) may be the result of four readings: 0.9979, 0.9978, 0.9976 and 0.9975 inch, all of which lie below the specified maximum manufacturing limit. In such a case, the quality control chart forecasts potential rejections instead of reporting actual ones, so that the manufacturer has time to make the adjustment in the process necessary to prevent rejections.

Use of Control Chart

To get and keep control of his production process, the manufacturer periodically takes a sample from the product while it is being made—for example, one sample every half hour, or one sample every day, or again, one sample of every lot of 1,000 as soon as completed. The size of the sample and the interval between two consecutive samples are to be decided on in each specific case.

Plotting the informative points on the control chart with the least possible delay enables the manufacturer to *take action* at once when quality appears to be disturbed. (For this reason, control limits thus used are also called *action limits*.) By such action the occurrence of defectives will often be prevented because the basic trouble in the process is adjusted before its effect has grown to the extent that the quality of the product exceeds its specified manufacturing limits.

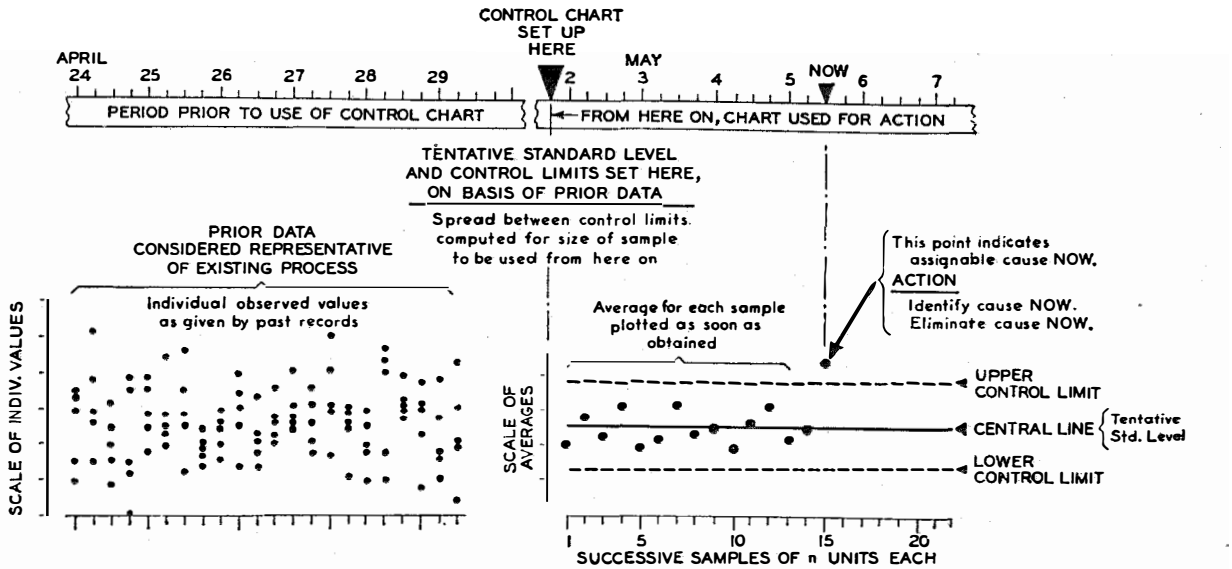

Fig. 3—How quality data from past records serve for starting a control chart used for controlling quality during production (control of process).

Starting a Process Control Chart

In Fig. 3 is shown how preparations for using the chart for control during production were made by taking 22 subgroups, each consisting of five observations, from the quality data recorded during the period from April 24 to 29 (left-hand side of Fig. 3). From these data, control limits were computed to be used as the initial set of limits for getting control of the process (right-hand side of Fig. 3). Starting on May 2, points were plotted from observations made on samples of n units each, taken four times a day from the product as soon as it came from the line. The first 14 points remained inside the control limits, but on May 5, point No. 15 fell outside the limits, thus calling for action.

Revision of Control Limits

After a certain number of points have been plotted on a chart used for process control, the limits should be computed again on the basis of the new observations. A decision should then be taken as to whether the original set of limits should be retained or revised. This check must be repeated from time to time and a compromise must be found between two requirements: the control limits must be checked often enough to be sure that they present a reliable picture of recent conditions in the process and, on the other hand, each new set of limits must be based on a number of observations large enough to be duly representative of collective quality. How this compromise is reached is a question of engineering judgment for which no detailed instructions can be given. However, the standard Control Chart Method of Controlling Quality During Production (Z1.3-1942) gives a series of illustrative examples which the user will find helpful in this and other respects, in solving his own problems.

Catch Trouble Early

The quality of any individual unit of product depends directly on the conditions existing in the process at the time that unit was made. Since the sooner we know that something has gone wrong with the process the better is our opportunity for adjusting it, the control chart should be kept at that point in the production process which is closest to the potential source of trouble. In one large plant manufacturing fire arms and ammunition, a control chart has been put on every production machine.

Other Advantages of Control

Wide practical experience has shown that once the assignable causes of variation have been weeded out and process control has been attained, there is no use in trying to keep variation in

quality within still closer limits. This knowledge is useful, for example, to the manufacturer who would like to change from selective fitting to assembly of interchangeable parts. This change requires a narrowing of the manufacturing limits. Without the statistical approach, the only way to find out if this can be done is by trial. If an effort in this direction is unsuccessful, the manufacturer has to decide whether or not he should keep on trying. If he does, and finally fails, he is back where he started from and has wasted time and money. The control chart method, however, enables him to decide in advance whether he can reduce the manufacturing tolerances sufficiently by adjusting the process in use or whether this can be done only by making basic changes in the process.

Maximum Reliability of Samples

When quality is controlled, data from samples of the product have the greatest possible reliability as a basis of judging current quality. This fact is important particularly where inspection by sampling is the only practical way of judging quality. Without control, there is a greater chance that sampling results will give a false picture of the quality of the product at hand, since each sample must be considered alone, rather than as one of a consistent series. This may easily lead to disagreement between the party supplying the product and the party receiving it —whether these parties be a seller and a buyer or two departments of the same manufacturing concern. Also, the great reliability of samples where control exists at a satisfactory level may make it possible to replace 100 percent inspection by sampling inspection, with consequent savings in time and cost.

Supplier-Purchaser Relations

An arrangement of great benefit to the supplier and the purchaser alike is one in which the supplier uses the control chart method for keeping control of his process and places his quality records at the disposal of the customer for the latter's information. The customer will then have evidence of the continuous reliability of the quality of the product supplied. He may safely forego rigid inspection of every shipment received, since he knows that the supplier, in making the product, has already used the most effective kind of quality control available. All the customer requires here is to take an occasional sample to make sure that the supplier has kept up his process control.

This supplier-purchaser relationship is of great value, for example, in subcontracting arrangements now widely used in war production work. Here, the control chart method not only helps the subcontractor reduce his percent defective, but also creates a better understanding between himself and the prime contractor.

Increasing Use of Control Chart

The U. S. War Department is using the control chart method extensively in the arsenals, and encourages its use by private industry. This technique is thus becoming an important factor in the war effort of this country and, also, in Great Britain, Australia and Canada, where the American War Standards in this field have been adopted by the respective national standardizing bodies.

It is recommended that industrial executives give the American War Standards on Quality Control their attention and have them applied to the repetitive processes used by their companies. Often the inspection data that are being collected in a manufacturing concern as a matter of routine for checking the quality of individual units of product are sufficient for the application of the control chart technique. These data can then be used to much better advantage, without increased cost of inspection. In some cases the cost of inspection may even be reduced. Therefore, the introduction of the new standards into practice should be most attractive to those in charge of industrial operations—especially at a time when this novel and refined method of quality control can make a major contribution to the winning of the war.

Antennae for Ultra-High Frequencies

Wide-Band Antennae

By LÉON BRILLOUIN

Columbia University

and

Consulting Engineer, Federal Telephone and Radio Laboratories, New York, N. Y.

*Editor's Note: This article is continued from "Electrical Communication," Volume 21, No. 4, 1944, and covers a general study of ultra-high frequency antenna problems as well as a comparison of the theories developed by different authors. Discussion of their contributions is illustrated by reproductions of figures from their published papers. Grateful acknowledgment is made for permission to utilize this material.**

10. Comparison with the Conventional Theory of Antennae. Radiation Resistance

The papers discussed in the preceding sections are either old ones (before 1910) or recent ones (after 1930). This is easy to explain inasmuch as the more elaborate and precise theory with which they deal is of special importance in ultra-short wave technique, and these waves have been mostly studied in the earlier times (Hertz and his followers, damped ultra-short waves) or recently when technical progress made the production of sustained ultra-short waves possible.

In the intervening period, interest centered largely on long waves on account of their very many practical applications. For such waves, an approximate theory of antennae was developed; we will call it the "conventional theory" as many engineers still use it in everyday work.

The conventional theory starts from a result obtained by M. Abraham for elongated ellipsoids: at resonance, the *current distribution* along the antenna wire is *practically sinusoidal*. This was taken as a fundamental assumption, and often explained by a comparison with similar formulae for waves propagating along a string or for sound waves in a pipe.

From the preceding discussion, this assumption may be checked. The drawings of Fig. 27 exhibit sinusoidal distribution at resonance, but a very different type of distribution below or above resonance. Similarly, Fig. 29 shows the current distribution along a thin ellipsoidal antenna; the curve $L/\lambda=0.5$ corresponds to resonance and is practically sinusoidal, while below resonance ($L/\lambda=0.6$) or above resonance ($L/\lambda=0.4$) the curves differ slightly from the sinusoidal. The change in the shape of the curve would be much larger for wider frequency intervals.

The ohmic resistance of the antenna wire is neglected, and electric waves along an infinite wire should propagate with the velocity $c=1/\sqrt{\epsilon_0\mu_0}$ (ϵ_0, dielectric power and μ_0, permeability of free space). The usual assumption is that the current distribution along the finite wire is sinusoidal;

$$I(x,t)=I_0\sin\frac{m\pi x}{l}e^{i\omega t},\quad m \text{ integer} \qquad (52)$$

where $I=0$ at both ends, $x=0$ and $x=l$, and ω is a complex quantity including radiation damping.† Assumption (52) would be valid if one could prove that the velocity of the waves remains constant and equal to c, even near the extremities, but this is very doubtful. If the damping is very small, ω is approximately

$$\omega=\frac{2\pi c}{\lambda}+i\alpha=\frac{m\pi c}{l}+i\alpha,\quad \lambda=\frac{2l}{m} \qquad (53)$$

with a small damping coefficient α, which can be obtained by calculation of the field radiated at great distance. This field is proportional to I_0, and the energy lost by radiation varies as I_0^2 and can be written as

$$W_r=RI_0^2 \qquad (54)$$

which results in the introduction of the "*radia-*

* See note at end of article.

† See, for instance, J. A. Stratton, Electromagnetic Theory, McGraw-Hill, 1941, p. 438.

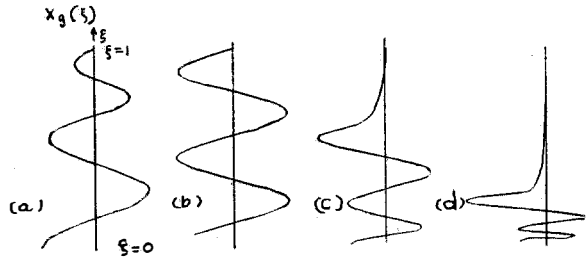

Fig. 27—Qualitative distribution of current in the ninth harmonic: (a) low frequency; (b) resonant frequency; (c) and (d) above resonance. (Reincluded from preceding installment for convenient reference.)

tion resistance" R playing a role similar to a resistance term; hence, the damping coefficient α can easily be obtained.

An alternative treatment of the problem was suggested by the author some years ago.[11] A resistance R, in the usual current theory, plays a double role: it appears in the RI^2 term for energy dissipation, and also as an electromotive force RI in the circuit. The author proved that this second aspect of the radiation resistance could be obtained if the field along the antenna were computed by the use of retarded potentials. The theory seemed applicable in practice, but S. Schelkunoff[12] recently advanced a troublesome and ingenuous argument: on the surface of a perfect metal, he noticed, the boundary condition requires that the electric field be normal to the surface. Hence, along a straight wire antenna, there can be no longitudinal component of the electric field and no radiation resistance RI term! Since a very important point is involved, let us examine the theory in some detail and see what sort of an approximation it represents. The method is the following (Stratton[7] p. 438, p. 455 or Brillouin[11]):

In the circuit of Fig. 30, call s a length measured along the circuit. For the sake of simplicity let us first neglect the diameter of the wire, an approximation that will be discussed subsequently. Let I_0 be the current measured by an ammeter, which should preferably be placed where the current is maximum (antinode). Call $i(s)$ the current at position s and $\sigma(s)$ the charge density at the same place. In attempting to apply the elementary formulae of circuit theory, the self induction L and the capacity C are defined as

$$\frac{1}{2}LI_0^2=\frac{\mu_0}{8\pi}\int\int(\vec{i}\cdot\vec{i}')\frac{dsds'}{r}$$

$$=\frac{\mu_0}{8\pi}\int\int ii'\cos\theta\frac{dsds'}{r}, \quad (55)$$

$$\frac{Q^2}{2C}=\frac{1}{8\pi\epsilon_0}\int\int\sigma\sigma'\frac{dsds'}{r},$$

where θ is the angle between the circuit elements ds and ds'.

These elementary formulae yield the well-known L and C coefficients of the circuit. They do not give anything like a resistance term. Here it must be recalled that this theory applies only at low frequencies and to circuits with dimensions extremely small compared to the wave length. As a matter of fact, the derivation of eq. (55) implies neglect of the very short time which is necessary for the propagation of the fields from ds' to ds. With this approximation, the electromotive forces induced by one circuit element ds' on another ds are exactly out of phase with the current. Their sum over the whole circuit yields the usual $L(di/dt)$ and $(1/C)\int idt$ terms.

But when high frequency oscillations are involved, these elementary calculations can not be used without some corrections. The field propagation from ds' to ds takes a small but finite time; hence, the electromotive force will no longer be exactly out of phase, but slightly advanced as shown on Fig. 31, and will yield components in phase with the current i. This is a very rough explanation of the origin of an electromotive force in phase with the current, which is the radiation resistance term.

A more precise calculation is possible by the use of "retarded potentials." When the current distribution $\vec{i}'ds'$ and the charge distribution

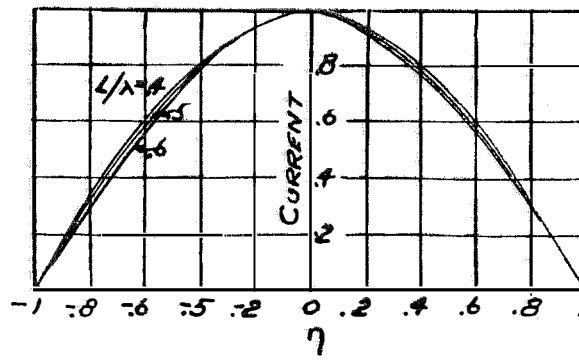

Fig. 29—Current distribution as function of η for thin ellipsoidal antenna.

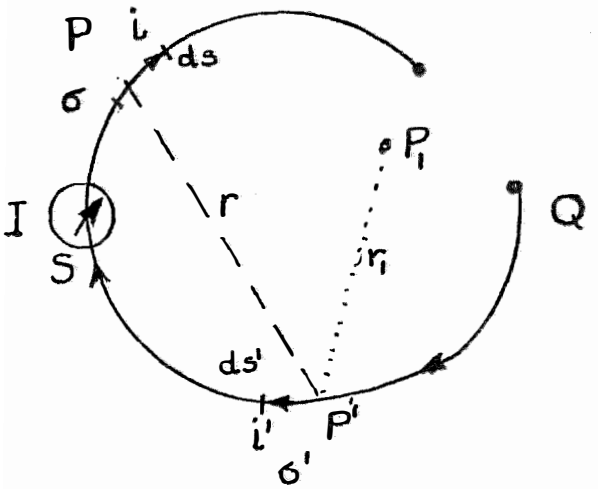

Fig. 30—Analysis of the field distribution near an electric circuit.

$\sigma' ds'$ are assumed, the retarded potentials at any point around the conductor are given by the formulae

$$\text{Scalar potential} \quad V(P_1) = \frac{1}{4\pi\epsilon_0}\int \frac{\sigma'^* ds'}{r_1},$$

$$\text{Vector potential} \quad \vec{F}(P_1) = \frac{\mu_0}{4\pi}\int \frac{\vec{i}'^* ds'}{r_1}. \qquad (56)$$

$$* \cdots t - \frac{r_1}{c}$$

The asterisks mean that σ' or i' must be taken at the time $t - r_1/c$ if one wishes to obtain the potentials V and $\vec{F}$ at P_1 at the time t. When these potentials have been obtained, the electric field h and the magnetic field H at P_1 are given by

$$\vec{h} = -\text{grad } V - \frac{\partial \vec{F}}{\partial t}$$

$$h_x = -\frac{\partial V}{\partial x} - \frac{\partial F_x}{\partial t} \quad \cdots \quad \cdots \qquad (57)$$

$$\mu_0 \vec{H} = \text{rot } \vec{F} \qquad \mu_0 H_x = \frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \quad \cdots \quad \cdots$$

These formulae apply for any arbitrary point P_1 in space; they can be used to compute the field at P, ds, and they give the rigorous expressions for the field, including the effect of finite velocity of propagation c, which is taken into account in eq. (56).

Neglecting the time delay due to propagation (as was done in (55)), the stars may be dropped from eq. (56) so that, for the whole circuit,

$$\text{Electrostatic energy} = \frac{1}{2}\int V(P)\sigma ds$$

$$= \frac{1}{8\pi\epsilon_0}\int\int \sigma\sigma' \frac{ds ds'}{r},$$

$$\text{Electromagnetic energy} = \frac{1}{2}\int (\vec{F}\cdot\vec{i}) ds \qquad (58)$$

$$= \frac{\mu_0}{8\pi}\int\int (\vec{i}\cdot\vec{i}') \frac{ds ds'}{r}$$

which is the same as (55). If, in addition, the current is assumed constant along the whole circuit, the self induction is

$$L = \frac{\mu_0}{4\pi}\int\int \cos\theta \frac{ds ds'}{r}, \quad I = i = i' \qquad (59)$$

a well-known formula due to Laplace.

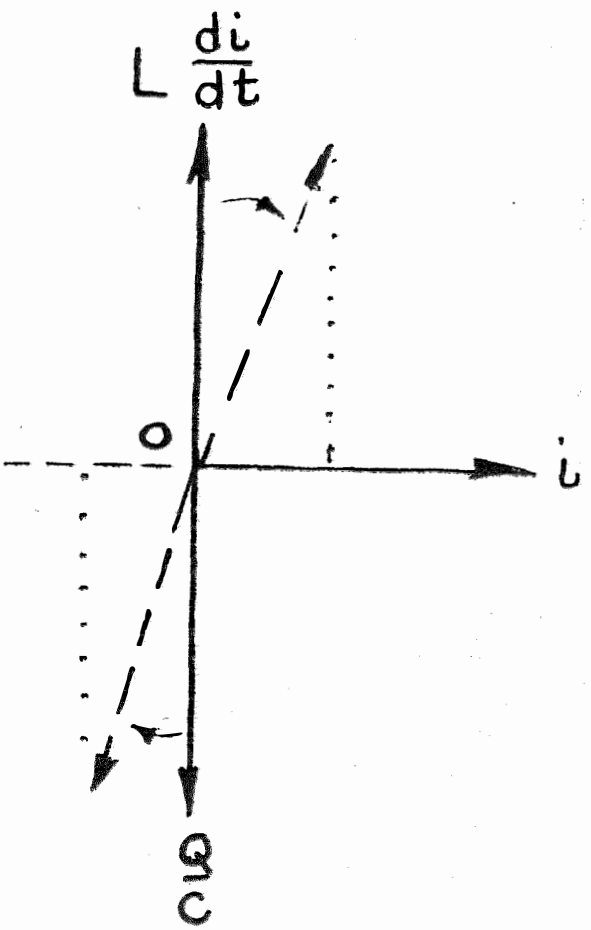

Fig. 31—Phase angle of the different terms.

In these first and very crude explanations, the finite radius of the wire was not considered. This proves a serious omission since r becomes zero when ds and ds' are just near each other and (eq. 59) consequently yields an infinite self induction.

Thus the first thing we learn is that it is necessary to make provision for the finite radius of the wire. The second is that the correct equations (56) (57) including the finite velocity of propagation c must be applied in order to derive the terms "radiation resistance."

11. Elementary Theory of the Rectilinear Antenna

Retarded potentials are especially useful when the charge and current distributions are known in advance since they immediately give the field distribution in the whole space according to Maxwell's equations. The same formulae also can be used when the electric charges and currents are not given "a priori" but must be determined afterwards on the basis of additional conditions.

Let us discuss a problem of the first type, representing a classical approach to the theory of a rectilinear antenna. From the general discussion of the preceding sections, it will be assumed that the current distribution, along the antenna wire, is practically sinusoidal when the antenna vibrates with almost sustained oscillations (including the case of free oscillations when the damping is very small) in the immediate neighborhood of resonance. If the antenna is a very fine wire of radius a and length l, extending from $x=0$ to $x=l$, the hypothesis of equation (52) yields

$$I=I(x)e^{i\omega t}=I_0 \sin\frac{m\pi x}{l}e^{i\omega t}, \quad m \text{ integer} \qquad (52)$$

ω is real for sustained oscillation; it may contain a small imaginary component in the case of damped oscillations.

As a general rule, when the current distribution is known, the charge density distribution results from the equation for conservation of electricity.

$$-\frac{\partial\sigma}{\partial t}=\frac{\partial I}{\partial x},$$

$$\sigma=\frac{i}{\omega}\frac{\partial I(x)}{\partial x}e^{i\omega t}=\frac{i}{\omega}\frac{m\pi}{l}I_0\cos\frac{m\pi x}{l}e^{i\omega t}. \qquad (60)$$

It should be noted that no account is taken of the capacity of the flat terminals at both ends of the wire, an approximation only justified for very fine wires when the ratio a/l can be neglected as compared with other terms of the order of magnitude of $1/\Omega$ (eq. 11). This corresponds to M. Abraham's approximations for very long and thin ellipsoids.

I and σ being known, the retarded potentials (56) may be formulated.

$$\sigma^*=\sigma\left(t-\frac{r}{c}\right)=\sigma e^{-i\omega(r/c)}, \quad I^*=Ie^{-i\omega(r/c)}, \qquad (61)$$

where r is the distance between two points located at x' and x along the antenna; $r=|x-x'|$.

The potentials at x are

$$V(x)=\frac{1}{4\pi\epsilon_0}\int_0^l\frac{\sigma^*(x')dx'}{r}$$

$$=\frac{i}{4\pi\epsilon_0\omega}e^{i\omega t}\int_0^l\frac{\partial I(x')}{\partial x'}\frac{e^{-i\omega(r/c)}}{r}dx', \qquad (62)$$

$$F_x(x)=\frac{\mu_0}{4\pi}e^{i\omega t}\int I(x')\frac{e^{-i\omega(r/c)}}{r}dx';$$

the vector potential has only an x component since the current everywhere flows in the x direction.

The electric field along the wire itself can now be computed according to (57). As a preliminary, the finite radius of the wire must be considered in order to avoid the infinity due to the $1/r$ factor. This is a rather delicate problem when discussed rigorously, but it finally amounts to the following:

$$\frac{1}{r}e^{-i\omega(r/c)} \quad \text{must be replaced by a function} \quad G_\omega(r) \qquad (63)$$

which behaves like $e^{-i\omega(r/c)}/r$ for large values of r and takes a very large value for $r=0$ in such a way as to yield finite integrals. This is a purely mathematical problem, discussed later;

$$V(x)=\frac{i}{4\pi\epsilon_0\omega}e^{i\omega t}\int_0^l\frac{\partial I}{\partial x'}G(r)dx',$$

$$F_x(x)=\frac{\mu_0}{4\pi}e^{i\omega t}\int_0^l I(x')G(r)dx', \qquad (64)$$

$$k^2=\frac{\omega^2}{c^2}=\omega^2\epsilon_0\mu_0=\frac{4\pi^2}{\lambda^2}.$$

Hence, by (57),

$$h_x(x) = -\frac{\partial V}{\partial x} - \frac{\partial F_x}{\partial t} = -\frac{ie^{i\omega t}}{4\pi\epsilon_0\omega}\int_0^l \left[\frac{\partial I}{\partial x'}\frac{\partial G}{\partial x} + k^2 I(x')G(r)\right]dx'. \quad (65a)$$

Here G is a function of $(x-x')$ as $r=|x-x'|$; consequently,

$$\frac{\partial G}{\partial x} = -\frac{\partial G}{\partial x'}. \quad (65b)$$

Integrating G by parts

$$h_x(x) = -\frac{ie^{i\omega t}}{4\pi\epsilon_0\omega}\left\{-\left|G(r)\frac{\partial I}{\partial x'}\right|_{x'=0}^{x'=l} + \int\left[\frac{\partial^2 I}{\partial x'^2}+k^2 I(x')\right]G(r)dx'.\right\} \quad (66)$$

According to the usual assumption (52), resonance on the m harmonic signifies that

$$\frac{m\pi}{l} = k = \frac{2\pi}{\lambda}, \quad \lambda = \frac{2l}{m}, \quad I(x') = I_0 \sin\frac{m\pi x'}{l}, \quad (67)$$

which cancels out the last integral. Thus we finally find a longitudinal electric field along the antenna:

$$h_x = \frac{iI_0 e^{i\omega t}}{4\pi\epsilon_0\omega}\frac{m\pi}{l}[(-1)^m G(r_2) - G(r_1)] \qquad \begin{pmatrix} r_2 = l - x \\ r_1 = x \end{pmatrix} \quad (68)$$

since

$$\frac{\partial I}{\partial x'} = I_0\frac{m\pi}{l}\cos\frac{m\pi x'}{l} = \begin{cases} I_0\dfrac{m\pi}{l} & \text{for } x'=0 \\ I_0\dfrac{m\pi}{l}(-1)^m & x'=l \end{cases}$$

This is exactly the formula given by Stratton [7] (P. 457 eq. 76a) but for a change of sign; Stratton uses $-i\omega t$ in the exponentials, hence $(-i)$ instead of $(+i)$ and the replacement of G by e^{-ikr}/r according to (63). The present calculation is much quicker than the one given by Stratton.

How can this formula be applied? Here is the point where we must consider the very fundamental remark of Schelkunoff: if the antenna wire is an ideal conductor, we know that there must be no tangential electric field along the wire. We can, however, use the result (68) in three different ways, the first of which is rigorous while the other two are only approximate:

I. In order to sustain continuous oscillations on the antenna under resonance conditions (67), it is necessary to add an external field h_x' which compensates the field (68). This external field may be produced by an external incident wave (of a very arbitrary type indeed).

The additional field h_x' gives to the antenna, per second, just as much energy W as is radiated by the antenna at great distance:

$$h_x' = -h_x, \quad W = \frac{1}{2}\mathbf{Re}\int_{x=0}^{x=l} h_x'\tilde{I}(x)dx = \frac{1}{2}RI_0^2, \quad (69)$$

where the sign $\sim$ means "imaginary conjugate" and **Re** indicates "real part of." The R term in this formula is the *radiation resistance* for the antenna at resonance on the *mth mode*. This result is perfectly rigorous, provided the h_x' field can be created along the antenna.

II. The field h_x' has a very peculiar distribution along the antenna, and to obtain exactly the desired field distribution would be difficult. What we can do is to use a field h_x'' giving to the antenna just the right amount W of energy per second, but with a different distribution along the wire:

$h_x'' \neq h_x'$,

$$W = \frac{1}{2}\mathbf{Re}\int h_x'\tilde{I}dx = \frac{1}{2}\mathbf{Re}\int h_x''\tilde{I}dx = \frac{1}{2}RI_0^2. \quad (70)$$

If this condition is fulfilled, the sustained oscillations ω are maintained in the antenna but, as a result of the difference between h'' and h', the current distribution will be slightly modified and may differ from the sinusoidal curve (52). It will be shown later that the change in current distribution is actually very small when the ratio a/l of radius to length of the antenna is small.

The above covers both the receiving antenna (case B §1) and the transmitting antenna (C §1).

III. How about the free oscillations (case A §1)? Here no external field is added, but an exponential decrease of the oscillations must be assumed. This means a complex

$$\omega = \omega_r + i\omega_i, \quad k = \frac{1}{c}(\omega_r + i\omega_i), \quad \omega_i \text{ small}. \quad (71)$$

The small ω_i correction is not sufficient to justify making the longitudinal field h_x of (66) exactly zero all along the antenna but this may be

realized as an average, so that the condition

$$W=\frac{1}{2}\mathbf{Re}\int h_x\tilde{I}dx=0 \qquad (72)$$

is satisfied, showing that the antenna does not receive any energy from outside. Condition (72) yields the damping term ω_i of free oscillations. Again, as in the preceding case, there will be a correction for the current distribution, but it will be very small for fine wires.

In order to complete the discussion, we must:
a—compute the radiation resistance R of (eq. 69)
b—justify the assumption made in Cases II and III regarding the smallness of the correction for current distribution.

Computation of the Radiation Resistance, from (69), yields

$$R=\mathbf{Re}\frac{-i}{4\pi\epsilon_0 c}\int_0^l \langle(-1)^m G(l-x) - G(x)\rangle \sin kx dx. \qquad (73)$$

Here it is permissible to revert from $G(r)$ to e^{-ikr}/r as in (63) since, in (73), the $1/r$ term is infinite only for $x=0$ or $x=l$ and, in both cases, $\sin kx$ is zero and the infinity is avoided. This means that the *radiation resistance does not depend on the radius of the wire*, at least for *fine wires* and at resonance, a very important conclusion inasmuch as other quantities do not show the same characteristic. Hence, in (eq. 73),

$$G(x)=\frac{\cos kx - i\sin kx}{x},$$

$$\mathbf{Re}(-i)(-1)^m G(l-x)=(-1)^{m+1}\frac{\sin k(l-x)}{l-x}=\frac{\sin kx}{l-x},$$

$$\mathbf{Re}(iG(x))=\frac{\sin kx}{x},$$

and

$$R=\frac{1}{4\pi\epsilon_0 c}\int_0^l\left(\frac{1}{l-x}+\frac{1}{x}\right)\sin^2 kx dx$$

$$=\frac{1}{4\pi\epsilon_0 c}\int_0^l\frac{2\sin^2 kx}{x}dx \qquad (74)$$

$$=\frac{1}{4\pi\epsilon_0 c}\int_0^l\frac{1-\cos 2kx}{x}dx,$$

which yields the well known "cosine integral" $Ci(2kl)$.

$$R=30(\log 2m\pi\gamma - Ci(2m\pi)) \text{ ohms}$$
$$2kl=2m\pi$$
$$\frac{1}{4\pi\epsilon_0 c}=30 \text{ ohms} \qquad (75)$$
$$\gamma=e^c=1.7811$$
$$C=\text{Euler-Mascheroni constant}$$

as in Stratton (P. 444, eq. 19–22 or P. 460 eq. 91). R represents the exact counterpart of the energy radiated at great distance.

Since the radiation resistance at resonance does not depend on the diameter of the wire, it is also independent of small changes in the shape of the antenna; for instance, the change from cylindrical to ellipsoidal shape. As a matter of fact, the radiation resistance *at resonance* for the successive harmonics m, as given in (eq. 75), yields results identical with those obtained by Ryder for ellipsoids (see eq. 32).

Fig. 32 shows this radiation resistance at resonance. The method outlined does not allow for the computation of radiation resistance between successive resonances; the theory is based on the assumption of sinusoidal current distribution (52), which holds only at resonance.

The radiation resistance computed for case I is certainly correct, but case I is a rather artificial one. In cases II and III, the result should be only approximate. An inadequacy involving these two cases remains: the field h_x as given by (68) becomes very large at both ends of the antenna when $x=0$, $x=l$ or $r_1=0$ or $r_2=0$. An exact computation even gives a logarithmic infinite value (see next section, eq. 81); further, the effect of both flat end surfaces has been neglected. These factors, however, drop out of the practical computation of the radiation resistance, but may alter materially the current and charge distribution near the ends of the antenna wire.

12. *A More Rigorous Theory—Comparison with a line of finite length or with a tuning circuit*

The success of this elementary theory must be considered as strictly limited to very fine and long antennae (ratio a/l negligible) operated at resonance. The reason it applies is that our first assumption (52) of sinusoidal current distribution automatically cancels the integral of eq. (66) and just one simple term remains.

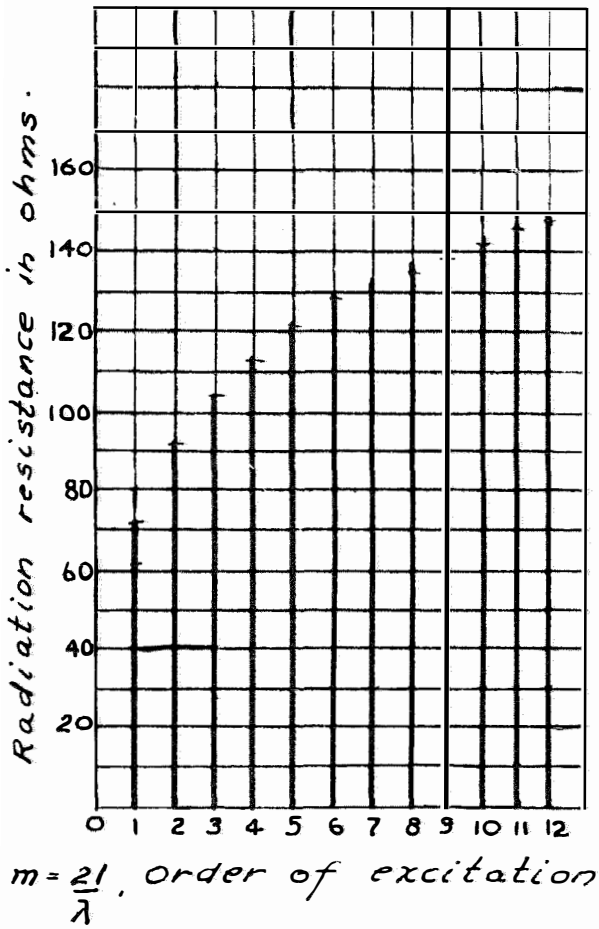

Fig. 32—Radiation resistance of a linear antenna at resonance.

The integral is the most important term in (66) because it contains the function $G(r)$ which becomes very large for $r=0$ $(x=x')$. If its two terms would not cancel each other, the result would be much larger than the term retained, and it would be impossible to compensate for it. Similarly, problems II and III could be discussed in the approximate manner of the preceding section since the sinusoidal current distribution is very stable as regards small perturbations.

In considering these general remarks and attempting to formulate a more satisfactory theoretical basis, we may proceed as follows: Disregarding any "a priori" assumption as to current distribution,

$$I(xt) = I(x)e^{i\omega t} \tag{76}$$

may be written instead of (52). The unknown function $I(x)$ is determined afterwards from the boundary conditions stated in each type of problem. We then proceed exactly as in equations (60) to (66) and again obtain (66):

$$h_x(x) = -\frac{ie^{i\omega t}}{4\pi\epsilon_0\omega}\left\{-\left|G(r)\frac{\partial I(x')}{\partial x'}\right|_{x'=0}^{x'=l} + \int_0^l\left[\frac{\partial^2 I}{\partial x'^2}+k^2 I(x')\right]G(r)dx'\right\} \tag{77a}$$

$$r=|x-x'|$$

giving the longitudinal field component along the antenna wire at point x. This h_x with the external field h_x' added to it (driving field h_x') must give zero on every point of the antenna since the antenna is assumed to be a perfect metal. Consequently, our fundamental and rigorous equation reads:

$$-h_x = h_x' \quad \text{driving external field.} \tag{77b}$$

These rigorous equations (77a and b) are satisfied in Case I of the preceding section (eq. 69). Cases II and III can be stated as follows:

II—a Receiving antenna, incident wave polarized parallel to the antenna

$$h_x' = Ee^{i\omega t}.$$

II—b Transmitting antenna:

$$h_x' = 0 \quad \text{for} \quad 0 \leqslant x \leqslant \frac{l-\epsilon}{2} \quad \text{or} \quad \frac{l+\epsilon}{2} \leqslant x \leqslant l$$

$$h_x' = \frac{V}{\epsilon}e^{i\omega t} \quad \text{for} \quad \frac{l-\epsilon}{2} < x < \frac{l+\epsilon}{2}$$

corresponding to an antenna energized at the middle point, across short strip ϵ, with driving potential V.

III—Free Oscillations

$$h_x' = 0 \quad \text{everywhere,}$$
$$\omega = \omega_r + i\omega_i \quad \text{complex.}$$

Remembering that $G(r)$ is very large for $r=0$, we note that in (77a) the integral is by far the most important term at every point x except $x=0$ and $x=l$. The method used in the preceding section amounts to the following: First, make the integral zero, necessitating sinusoidal current distribution (52) since $\partial^2 I/\partial x'^2 + k^2 I$ then becomes zero. Second, treat the remaining term (68) as a small perturbation, which will slightly

modify the current distribution, and result in an average as a resistance term, the "radiation resistance" as computed in (75).

Let us now try to compare the fundamental equation (77) with the conventional line equations. We first rewrite (77) as follows:

$$\frac{1}{4\pi\epsilon_0}\int_0^l \frac{\partial^2 I}{\partial x'^2}G(r)dx' + \omega^2\frac{\mu_0}{4\pi}\int_0^l I(x')G(r)dx' = \frac{1}{4\pi\epsilon_0}\left| G(r)\frac{\partial I}{\partial x'}\right|_0^l - i\omega h'_x e^{-i\omega t}. \quad (78a)$$

This may be compared with wave propagation along a conventional line, with Ldx and Cdx as self induction and capacity per length dx;

$$\frac{1}{C}\frac{\partial^2 I}{\partial x^2} + \omega^2 LI(x) = E(x), \qquad v = \frac{1}{\sqrt{LC}}(\text{wave velocity}) \quad (78b)$$

where $E(x)$ represents an external field applied along the line at x.

Comparing equations (78a) and (78b) term by term, we obtain a physical picture of the meaning of (77). This picture should not be unfavorable when the radius a of the wire is very small, so that $G(r)$ exhibits a very high maximum at $r=0$. Both integrals of (78a) could then be approximated as

$$\frac{1}{C}\approx\frac{B}{4\pi\epsilon_0}, \quad L\approx\frac{\mu_0 B}{4\pi}, \quad B=\int_0^l G dx', \quad (79)$$

with

$$\int_0^l f(x')G(|x-x'|)dx' \approx Bf(x).$$

The very strong maximum of G for $x=x'$ practically permits this approximation. Thus the "equivalent conventional line" would exhibit very large L and $1/C$ coefficients, justifying the qualitative explanation given at the end of the preceding section.

This comparison is, however, not very satisfactory and should not be used for more than a rather crude visualization of the real facts. Many authors have been misled by attempts at treating antennae as some sorts of lines, and such procedures are hardly to be recommended. A better comparison could probably be achieved by computing the self induction L_0 and the capacity C_0 of an equivalent resonant circuit by the relations (similar to 58) (64)

$$\frac{1}{2}L_0 I_0^2 = \frac{1}{2}\mathbf{Re}\int F(x)\bar{I}(x)dx = \frac{\mu_0}{8\pi}\mathbf{Re}\int\int I(x')G(r)\bar{I}(x)dxdx',$$
$$\frac{1}{2}\frac{I_0^2}{C_0\omega^2} = \frac{1}{2}\mathbf{Re}\int V\bar{\sigma}dx = \frac{1}{8\pi\epsilon_0\omega^2}\int\int\frac{\partial I'}{\partial x'}G\frac{\partial\bar{I}}{\partial x}dxdx'. \quad (80)$$

The smaller the wire, the larger G and the self induction L_0; the radiation resistance remains practically constant (eq. 73–75). Hence the damping decreases as $R/2L_0$, which is exactly the result obtained by M. Abraham for elongated ellipsoids. This may also be explained as follows: For fine antennae, the amount of energy radiated, per second, at great distance, is practically independent of the radius of the wire. But when the radius is very small, there is a large amount of electromagnetic energy accumulated in the field, just near the wire. The ratio of energy radiated to energy stored in the local field is low, which means small damping.

All this discussion shows the very limited field of application of the elementary discussion summarized in §11. It must be considered as a first approximation, valid only for extremely fine wires, used at one of their resonance vibrations.

In order to complete these remarks, let us indicate briefly how the function $G(r)$ of (63) can be obtained. Its expression would be difficult to compute for low frequencies where the skin effect is not very marked. At very high frequencies, one can assume that the current flows only along the surface of the cylindrical wire, and $G(r)$ results from an averaging [13] over two cross

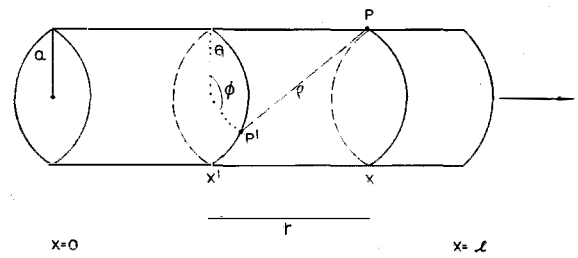

Fig. 33—Interaction between two points P—P' on a cylindrical antenna.

sections located at x' and $x=x'+r$ (Fig. 33). This yields

$$\rho = PP'$$

$$G(r)=\frac{1}{2\pi}\int_0^{2\pi}\frac{e^{-ik\rho}}{\rho}d\varphi,\quad \rho^2=r^2+2a^2(1-\cos\varphi). \quad (81)$$

Approximate expressions for G can be worked out for very fine wires; they show that G is logarithmically infinite for $r=0$, a circumstance that removes the difficulties in the computation of the integrals * but may still prove troublesome as to the role played by both end-surfaces (see end of section 11).

Let us again consider the fundamental equation (78a). The first term on the right hand side is the same as (68) if, for the first approximation, the sinusoidal current distribution (52) is used for an antenna oscillating on its mth mode of vibration. This correcting term plays the role of an additional series impedance $Z'(x)$ at each point x of the antenna

$$Z'(x)=\frac{-i}{4\pi\epsilon_0\omega I(x)}\left|G(r)\frac{\partial I(x')}{\partial x'}\right|_{x'=0}^{x'=l}$$

$$=\frac{-i}{4\pi\epsilon_0 c\sin kx}[(-1)^m G(l-x)-G(x)]. \quad (82)$$

Use is made of eq. (68) for the actual computation. Except in the immediate neighborhood of both end-faces ($x=0$ or $x=l$), G may be replaced by

$$G(r)\approx\frac{e^{-ikr}}{e}=\frac{\cos kr-i\sin kr}{r}$$

and

$$Z'(x)=\frac{1}{4\pi\epsilon_0 c}\left[(-1)^{m+1}\frac{i\cos k(l-x)+\sin k(l-x)}{(l-x)\sin kx}+\frac{i\cos kx+\sin kx}{x\sin kx}\right].$$

Since for the mth mode of vibration

$$(-1)^{m+1}\sin k(l-x)=\sin kx \qquad kl=m\pi$$
$$(-1)^m\cos k(l-x)=\cos kx$$

$$Z'(x)=\frac{1}{4\pi\epsilon_0 c}\left[\frac{1}{l-x}+\frac{1}{x}+i\operatorname{cotg} kx\cdot\left(\frac{1}{x}-\frac{1}{l-x}\right)\right]$$

$$=R'(x)+iJ'(x) \quad (83)$$

* As is well known: $\int_0^{r_0}\log r\,dr=r_0(\log r_0-1)$ is finite.

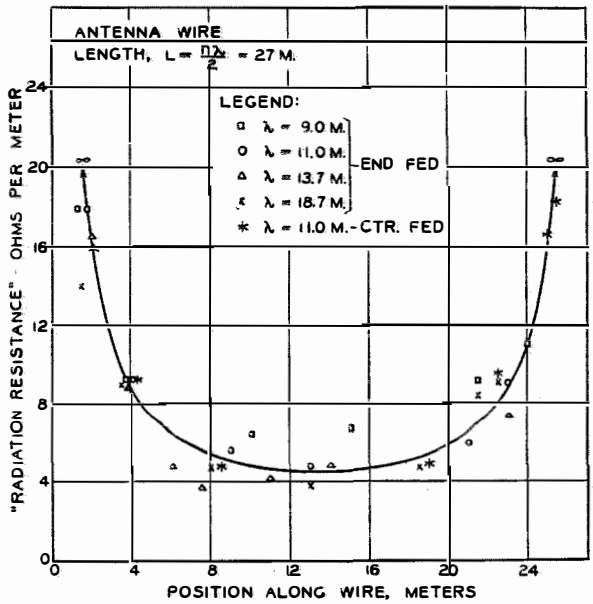

Fig. 34—Distribution of radiation resistance along an antenna. Antenna Wire Length $L\frac{n\lambda}{2}=27M$.

Legend:

□ λ = 9.0 M
○ λ = 11.0 M
△ λ = 13.7 M
× λ = 18.7 M
} END FED

*λ = 11.0 M − CTR FED

which may be interpreted as yielding the distribution of radiation resistance R' and radiation reactance J' along the antenna:

$$R'=30\left(\frac{1}{l-x}+\frac{1}{x}\right)=\frac{30l}{x(l-x)}\ \text{ohms},$$
$$J'=30\operatorname{cotg} kx\cdot\frac{l-2x}{x(l-x)}. \quad (84)$$

These formulae were obtained by Pistolkors [14] and tested experimentally by Schelkunoff and Feldman.[15] They check rather well with experimental results. It must, however, be noted that both R' and J' are infinite at both ends of the antenna. This proves once more the importance of the role played by both end-surfaces, a point mostly overlooked even by very recent authors.

As an example of experimental results from radiation resistance formulae, the diagram of Fig. 34 shows how the experimental points fall in the neighborhood of the theoretical curve. The figure is taken from a paper by S. A. Schelkunoff and C. B. Feldman [15] that also contains an interesting theoretical discussion.

13. *Systematization of Methods of Retarded Potentials*

The preceding section shows definitely how a rigorous theory of the cylindrical antenna could be worked out. Starting with an arbitrary current distribution along the antenna (76), the field in the whole space is computed by the use of retarded potentials. Then one takes the field on the surface of the wire and formulates the necessary boundary conditions: for ideal metal, the tangential field must be zero; for actual metal, the well known superficial impedance may be used (see for instance Schelkunoff,[15] eq. (16)).* This yields an integrodifferential equation the solution of which is necessary for finding the actual current distribution $I(x)$ along the antenna, introduced at the beginning as an unknown function (76). In addition, the external driving fields should be taken into account in the case of receiving or transmitting antenna, as explained after eq. (77b).

The method was proposed long ago [17–20] and used by many authors. Some fundamental difficulties, however, require explanation.

In the preceding sections, no exact consideration was given to the *role played by the end surfaces.* It was simply assumed (eq. 52) that the current was zero at both ends of the wire. This is too crude an approximation, as can be readily understood.

An actual solid cylindrical antenna must be terminated, at both ends, either by flat cross sections or by some sort of rounded surfaces (Fig. 35). These end surfaces exhibit a certain capacity Γ which is, roughly speaking, proportional to the radius a of the cylinder;

$$\Gamma = Ka \tag{85}$$

the coefficient K depending on the exact shape of the end-surface. This end-capacity takes a charge

$$Q = \Gamma V = KaV \tag{85a}$$

where V = potential of the end-surface and plays a two-fold role. First, it needs a non-vanishing current at the end of the cylindrical wire:

$$I = \pm\frac{\partial Q}{\partial t} = \pm i\omega Q = \pm i\omega KaV \quad \begin{cases} +, x=l \\ -, x=0 \end{cases} \tag{86}$$

Second, the charge Q on the end-faces necessitates a corrective term in our field equation. Assuming that one can, as a first approximation, ignore small currents along the rounded end-surfaces and compute the potential as if the charge were concentrated at the middle points P_0, P_l, the additional retarded potential at a point P would be

$$V(P) = \frac{Q_0}{4\pi\epsilon_0}\frac{e^{-ikr_1}}{r_1} + \frac{Q_l}{4\pi\epsilon_0}\frac{e^{-ikr_2}}{r_2} = \frac{KaV_0}{4\pi\epsilon_0}\left[\frac{e^{-ikr_1}}{r_1} + (-1)^m\frac{e^{-ikr_2}}{r_2}\right] \tag{87}$$

mode m

$$Q_l = (-1)^m Q_0$$

$$V_l = (-1)^m V_0$$

V_0 is the potential at $x=0$. This additional contribution to the scalar potential at P results in a further term for the electric field. What we are interested in is the electric field h_x along the surface of the wire. We have already computed in (77a) the most important part of the field, which is due to the charges and currents on the wire. If we now add the contribution of $V(p)$,

$$h_x = \frac{-ie^{i\omega t}}{4\pi\epsilon_0\omega}\left\{-\left|G(r)\frac{\partial I}{\partial x'}\right|_0^l + \int_0^l\left[\frac{\partial^2 I}{\partial x'^2} + k^2 I\right]G\,dx'\right\} - \frac{KaV_0}{4\pi\epsilon_0}\frac{\partial}{\partial x}\left[\frac{e^{-ikr_1}}{r_1} + (-1)^m\frac{e^{-ikr_2}}{r_2}\right] \tag{88a}$$

$$r_1^2 = x^2 + a^2$$

$$r_2^2 = (l-x)^2 + a^2$$

the fundamental equation corresponding to (77b) is

$$-h_x = h_x' \quad \text{driving field.} \tag{88b}$$

Hence this attempt shows the need for two corrections: eq. (86) instead of $I=0$ on the end-faces, and the additional term in (88a).

This treatment, however, is not quite rigorous as the computation of the coefficient K in (85) does not appear very easy, and the assumption made in deriving (87) regarding charge concen-

* Page 522. The boundary conditions for an actual metal are:

$h_x = zI$, $z = \frac{1}{2\pi a}\sqrt{\frac{i\omega\mu}{g}}$, z surface impedance of the wire, $I(x)$ total current, $g\cdot\mu$ conductivity and permeability.

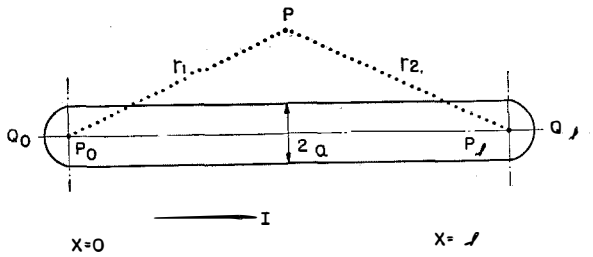

Fig. 35—Action of the end surfaces, and field created at point P.

tration is nothing but a rough approximation. More comprehensive rationalization along these general lines, however, has been attempted (see 13, 11).

The best way to indicate the importance of these corrections is to take some examples where they are not needed, and to show that these problems exhibit very exceptional features.

Let us assume, for instance, that the *wire* is so *fine* that the end capacity can reasonably be neglected:

$$a\to 0, \quad \Gamma\to 0$$

The corrective term in (88a) drops out, and the condition at the end is again $I=0$. As another example, consider a hollow cylinder with finite a but no end-surfaces, which should make $K=0$, at least at the limit of infinitely thin cylinder walls.

In both cases, the complicated $G(r)$ functions differ materially but may be eliminated by suitable transformation, thus enabling us to discuss both problems at once. Assuming no external driving field, i.e., free vibrations of these antennae, (88a) reduces to (77a) and yields

$$\int_0^l \left[\frac{\partial^2 I}{\partial x'^2}+k^2 I(x')\right] G(|x-x'|)dx' - G(l-x)\left(\frac{\partial I}{\partial x}\right)_l + G(x)\left(\frac{\partial I}{\partial x}\right)_0 = 0,$$

$$I=0 \quad \text{for} \quad x=0 \quad \text{or} \quad x=l \tag{89}$$

$$\left(\frac{\partial I}{\partial x}\right)_l = (-1)^m \left(\frac{\partial I}{\partial x}\right)_0 \quad m\text{th mode of vibration.}$$

Here we introduce a function δ of variable ξ with the definitions:

$$\delta(\xi)=0 \quad \text{for} \quad \xi<0 \quad \text{or} \quad \xi>\alpha$$

$$\delta(\xi)=\frac{1}{\alpha} \qquad 0\leqslant\xi\leqslant\alpha \tag{90}$$

If α is very small (or even if $\alpha\to 0$), then the following property for any arbitrary function $f(x)$ which exhibits only slow variation obtains:

$$\int_{-\infty}^{\infty} f(\xi)\delta(\xi-x)d\xi = \frac{1}{\alpha}\int_{\xi=x}^{\xi=x+\alpha} f(\xi)d\xi = \frac{1}{\alpha}\overline{f(x)}\alpha \to f(x) \quad \text{if} \quad \alpha\to 0 \tag{91}$$

With the help of this δ function, integral equation (89) becomes

$$\int_0^l G(|x-x'|)\left[\frac{\partial^2 I}{\partial x'^2}+k^2 I(x')+\left(\frac{\partial I}{\partial x}\right)_0 \delta(x') + (-1)^{m+1}\left(\frac{\partial I}{\partial x}\right)_0 \delta(l-x')\right]dx'=0 \tag{92}$$

since, for instance,

$$\int G(|x-x'|)\delta(x')dx' = G(|x-0|) = G(x).$$

This integral equation (92) is obviously solved for all values of x provided the bracketed portion can be made equal to zero.

$$\frac{\partial^2 I(x')}{\partial x'^2}+k^2 I(x') + \left(\frac{\partial I}{\partial x}\right)_0 [\delta(x')+(-1)^{m+1}\delta(l-x')]=0 \tag{93}$$

is a differential equation which is not very hard to analyze. The term with the δ functions is zero all along the wire but for two very short sections at both ends, namely,

$$[\delta(x')+(-1)^{m+1}\delta(l-x')] = \begin{cases} \frac{1}{\alpha} & 0\leqslant x'\leqslant\alpha \quad \text{I} \\ 0 & \alpha<x'<l-\alpha \quad \text{II} \\ (-1)^{m+1}\frac{1}{\alpha} & l-\alpha\leqslant x'\leqslant l \quad \text{III} \end{cases} \tag{94}$$

The problem now involves using a sinusoidal current distribution in the interval II and joining

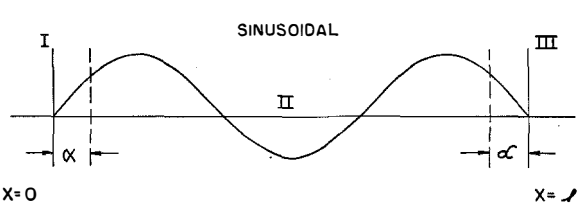

Fig. 36—Current distribution along the antenna (very fine wire or hollow cylinder).

it with the solutions obtained in the two small sections I, III, at both ends. The solution is easily obtained; the resulting current distribution is shown approximately in Fig. 36. In the short sections I, III, the curve is slightly perturbed; at the limit $\alpha \to 0$ this perturbation vanishes completely. Strangely enough, the bracket with the two δ functions in (93) thus plays an insignificant role, and the solution is exactly the elementary one

$$I = I_0 \sin \frac{m\pi x}{l}, \quad k = \frac{m\pi}{l} = \frac{\omega}{c}, \tag{95}$$

as in (52). Hence, the problem of free oscillation yields a real ω value with no imaginary term, i.e., *no damping*.

This is a well known result (M. Abraham, 1898!) for the infinitely fine antenna, the physical meaning of which has been discussed in §12 following eq. (80). For the case of the hollow cylinder, the explanation should be similar. The radiation resistance is still finite; hence the energy radiated per second is finite, but there must be an infinite amount of energy stored near the hollow cylinder because of the infinite electric field on the sharp edges of both terminals. The second problem is certainly of less practical importance since realization of such an antenna would mean building a hollow pipe thinner than the skin-effect layer!

14. *Methods of Oseen and Hallen—General Analysis*

The methods of Oseen [17] and Hallen [18] are practically the one of §§12, 13, but these authors utilize a different G function definition which must be discussed very carefully in connection with the difficulties emphasized in the preceding section concerning the importance of the role played by the end surfaces.

In order to make provision for the finite radius of the wire, we introduced in (63) a function G, which represents the average of $(1/\rho)e^{-ik\rho}$ for points taken all around the cross section at x' (where the current actually flows); the potentials were considered at a point x with $r = |x - x'|$. In deriving the G formula (81) (Figure 33), two essential facts were taken into account:

I —For ideal metal, currents and charges are located on the surface of the conductor (at x');

II—What we need is the field on another point of the surface of the conductor (at x), in order to substantiate the statement that this field has no tangential component (ideal metal).

The function G, obtained in this way, is a rather intricate one, hard to manipulate numerically. Oseen and Hallen simplify the problem by making two important assumptions:

$$\left.\begin{aligned} &\text{A}—G = \frac{e^{-ikr'}}{r'}, \quad r'^2 = (x-x')^2 + a^2 = r^2 + a^2, \\ &\text{B}—I = 0 \quad \text{for} \quad x = 0 \quad \text{and} \quad x = l. \end{aligned}\right\} \tag{96}$$

These simplifications will be discussed together. As will be seen, the error introduced by B is partly compensated by A, a very curious coincidence which must be clearly understood.

The difficulties connected with assumption B were explained in the preceding section. Let us

---

Note

Let us consider the first interval $0 \leqslant x' \leqslant \alpha$ where solution is required of an equation

$$\frac{\partial^2 I}{\partial x'^2} + k^2 I = \frac{D}{\alpha}, \quad D = \left(\frac{\partial I}{\partial x'}\right)_{x'=0}.$$

The solution is obviously a constant $D/k^2\alpha$ plus a sinusoidal function of kx' which may be written

$$I(x') = \frac{D}{k^2\alpha} - \frac{D}{k \sin \psi} \cos (kx' + \psi), \quad \frac{\partial I}{\partial x'} = D \frac{\sin (kx' + \psi)}{\sin \psi}.$$

$$\left(\frac{\partial I}{\partial x'}\right)_{x'=0} = D,$$

as it should be,

I must be zero for $x' = 0$;

hence,

$$\tan \psi = k\alpha, \quad \psi \to 0 \text{ as } \alpha \to 0,$$

and

$$\frac{Ik}{D} = \frac{\cos \psi - \cos (kx' + \psi)}{\sin \psi} \approx \frac{(kx' + \psi)^2 - \psi^2}{2\psi}$$

$$= \frac{k^2 x'^2}{2\psi} + kx' \approx k\left(\frac{x'^2}{2\alpha} + x'\right)$$

after expanding the cosines of ψ and $kx' + \psi$, both small quantities. This shows that at the end of the interval I is still small and $\partial I/\partial x$ has doubled.

$$\begin{aligned} &x' = 0, \quad x' = \alpha, \\ &I = 0, \quad I = \frac{3}{2} D\alpha, \qquad I = D\left(\frac{x'^2}{2\alpha} + x'\right), \\ &\frac{\partial I}{\partial x'} = D, \quad \frac{\partial I}{\partial x'} = 2D, \qquad \frac{\partial I}{\partial x'} = D\left(\frac{x'}{\alpha} + 1\right). \end{aligned}$$

These results justify the statements above. A similar treatment applies to the other terminal near $x' = l$.

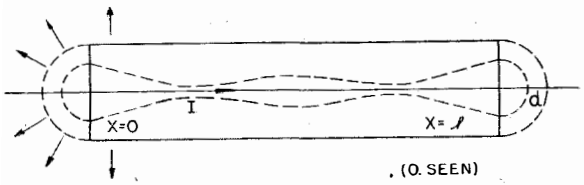

Fig. 37—Field distribution and end surfaces, assuming Oseen's viewpoint.

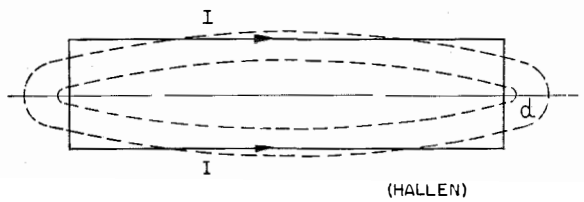

Fig. 38—Field distribution and shape of the antenna, according to Hallen's viewpoint.

now examine the approximation A. *Oseen* takes the following point of view:

Instead of considering the actual current, which flows along the surface of the wire, he computes an "equivalent fictitious current" which is supposed to flow along the axis of the cylindrical wire. He consequently justifies formula (96-A) by the assumptions:

$$\left.\begin{array}{l}\text{Charges and currents along the axis, at } x' \\ \text{Field observed on the cylinder of radius } a, \\ \quad \text{at } x.\end{array}\right\} \quad (97)$$

Obviously, the "equivalent" currents and charges along the axis may differ materially from the actual currents and charges on the cylindrical surface. Further, there is the fundamental question, whether the electromagnetic field, created around the antenna by the actual superficial currents, can always be obtained by a convenient choice of fictitious axial currents. This question remains open. It probably can be answered affirmatively for fine wires, but not for sizable wires.

Hallen assumes a different viewpoint, and works with the actual superficial currents, but computes the field on the axis rather than on the surface of the cylinder:

$$\begin{array}{l}\text{Charges and currents on the surface of the} \\ \quad \text{cylinder of radius } a, \text{ at } x' \\ \text{Field computed on the axis of the cylinder,} \\ \quad \text{at } x.\end{array} \quad (98)$$

Here the approximation is clear. When the longitudinal field is zero on the axis of symmetry, it is still very small (of the order of magnitude a^2) on the cylindrical surface a of the wire. Hence the rigorous boundary condition for the field has been replaced by an approximation, which should hold good for fine wires.

In order to visualize the difference between these two interpretations, the qualitative shape of surfaces, orthogonal to the electric lines of forces, has been sketched in Figs. 37 and 38 for the Oseen and Hallen cases. With the Oseen method, the complete metallic surface is an exact cylinder of radius a and length l, completed by two rounded ends, the precise shape of which could be obtained from drawings of the field distribution. The shape of the end-surfaces would be different for each mode m of oscillation. Following the Hallen procedure, the exact shape of the conductor resembles a sort of long ellipsoid. In both cases, the total length is $l+2d$, d being the length of each rounded end, a quantity of the order of magnitude of a.

All the practical applications of Hallen's method have been computed for the case of fine wires, where the ratio a/l can be neglected and the method is valid. It should be emphasized, however, that it applies only to *solid cylindrical* wires, with *rounded end-surfaces.* One should not, therefore, be surprised that the results appear very different from those obtained for the hollow pipe, discussed in the preceding section.

It is also to be anticipated that an exact treatment of the solid cylinder with *flat ends* might yield different results. This case has *never been rigorously discussed.* Such a structure possesses sharp edges at the border of the two terminal cross sections, and may very well exhibit difficulties similar to the hollow cylinder. An attempt at a rigorous theory [13] shows that the problem is difficult. If one tries to use the Oseen-Hallen methods for thick cylinders, the problem arises of how to improve the results, but Oseen's interpretation should be easier. One should first draw the exact shape of the rounded end-surfaces, then it would not be difficult to calculate the actual current along the cylinder. From the magnetic field H_a on the cylindrical surface, the current density i immediately results ($H_a=i$,

rational units) and the total current

$$I = 2\pi a i = 2\pi a H_a \qquad (99)$$

It should be emphasized that the second condition (96B) could probably not be maintained. In order to satisfy exactly the boundary condition along the cylinder $0 \leqslant x \leqslant l$, currents may be required along the axis, from $-d$ to $l+d$. This is suggested by Fig. 37.

Both Hallen and Oseen discussed possible applications to actual metals (of finite conductivity) and to different shapes of variable cross section. Due to the difficulties still remaining in the preceding problem, such extensions appear rather premature.

15. *Methods Available for Solving the Integral Equations*

Once the fundamental assumptions of Oseen and Hallen are accepted, the problem is to find practical methods for the resolution of the integral equation, which is the same as (77a and b) but with (96) for the function $G(r)$:

$$\int_0^l \left[\frac{\partial^2 I}{\partial x'^2} + k^2 I(x')\right] G(r)\,dx' - \left(\frac{\partial I}{\partial x'}\right)_{x'=l} G(l-x) + \left(\frac{\partial I}{\partial x'}\right)_{x'=0} G(x) = -i4\pi\epsilon_0\omega h_x'(x) \qquad (100)$$

where $k = \omega/c$ and $h_x'(x \cdot t) = h_x'(x)e^{i\omega t}$ is assumed for the driving field h_x' applied to the antenna. This is practically the same equation as (89) but for the right hand term. The results, however, appear very different, because of the assumptions (96A and B) discussed in the preceding sections.

The solution of eq. (100) can be obtained by different methods, all of which should be used and carefully compared as they rely on different systems of approximations.

A method of successive approximation has been proposed by the author [13] but not yet tried for actual computation. It would probably yield the solution as a Fourier series expansion:

$$I(x') = \sum_n A_n \sin k_n x', \quad k_n = \frac{n\pi}{l}. \qquad (101)$$

Another method consists in starting from a Fourier expansion (101) and substituting directly in the integral equation (100). This second method has been used by Francis Perrin * for systematic computations. It gives good results, leads to mathematical problems of standard type and well established methods of approximation. Substituting (101) in (100) and performing the x' integration, one first obtains an expansion

$$\sum_n A_n[\Phi_n(x) + (-1)^{n+1}k_n G(l-x) + k_n G(x)] = -i4\pi\epsilon_0\omega h_x'(x). \qquad (102)$$

The A_n coefficient is of course complex, and the integral Φ_n can be expressed in terms of integral sines and cosines. The integration proceeds along the same lines as (75) or (82). The next step is to multiply both sides by $e^{ik_m x}$ and to integrate again from $x=0$ to $x=l$, a procedure which simply amounts to analyzing the driving field h_x' in Fourier series terms. This yields an infinite system of simultaneous linear equations of the standard type

$$\sum_n A_n H_{nm} = D_m,$$

$$H_{nm} = \int_0^l [\Phi_n + (-1)^{n+1}k_n G(l-x) + k_n G(x)]e^{ik_m x}dx, \qquad (103)$$

$$D_m = -i4\pi\epsilon_0\omega \int_0^l h_x' e^{ik_m x}dx.$$

The matrix coefficients H_{nm} and the right hand terms D_m must be computed for each model of antenna. A practical solution of eq. (103) can be obtained by keeping only the first few terms of the expansions. One may, for instance, take

$$n \leqslant 5, \quad m \leqslant 5$$

reducing the computations to ten numerical coefficients since the A_n terms represent complex numbers. Inasmuch as equations of type (103) occur very frequently in problems of applied mathematics, numerical methods of solution have been very carefully considered.

The method followed by Hallen and R. King [21] is different. It can be better explained by first noticing some interesting facts involving the behavior of the potentials on the surface of cylindrical antennae. There are only two components for the potential: the scalar potential V and the F_x component of the vector potential

* A paper written in 1940 and not published.

(see eq. 62). F_x satisfies a usual wave equation

$$\Delta F_x - \frac{1}{c^2}\frac{\partial^2 F_x}{\partial t^2} = \Delta F_x + k^2 F_x = 0, \quad k = \frac{\omega}{c}$$

and its basic validity makes it applicable to the formulae for retarded potentials, derived above. Using cylindrical coordinates ρ, θ, x,

$$\frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\frac{\rho \partial F_x}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 F_x}{\partial \theta^2} + \frac{\partial^2 F_x}{\partial x^2} + k^2 F_x = 0. \quad (104)$$

In seeking symmetrical solutions, which do not contain θ, the second term is zero. Turning now to the first term, we wish to prove that it is also zero. This will be easier if we assume the point of view of Oseen [(96) and (97)]. Referring to eq. (57), in cylindrical coordinates, we notice that

$$\mu_0 H_\theta = -\frac{\partial F_x}{\partial \rho}.$$

Further, $2\pi\rho H_\theta$ is the contour integral of H_θ along a circle of radius ρ around the antenna. This is, according to Maxwell's equation, equal to the flux, through this circular area, of $J + \partial D/\partial t$, J being the current density and $\partial D/\partial t$ the displacement current. But, near the surface of the conducting cylinder, the electric field is radial; hence the flux of $\partial D/\partial t$ is naught. So, when ρ very nearly equals a (radius of the antenna), the flux is constant and independent of ρ. Hence $\mu_0 \rho H_\theta = -\rho(\partial F_x/\partial \rho)$ is a constant, and the first term in (104) drops out. This means that, along the surface of the cylindrical antenna, eq. (104) reduces to

$$\frac{\partial^2 F_x}{\partial x^2} + k^2 F_x = 0, \quad \rho = a, \quad (105)$$

which proves that F_x is an exact sinusoidal function. The same is true of the scalar potential V on account of the well-known Lorentz condition

$$\text{div } F = -\mu_0 \epsilon_0 \frac{\partial V}{\partial t} \quad \text{or} \quad \frac{\partial F_x}{\partial x} = -i\frac{k}{c}V. \quad (106)$$

We thus obtain, from Oseen's approach, the very important result: *the current distribution is not sinusoidal, but the potential distribution is always exactly sinusoidal along the cylindrical antenna.* The same result would apply, following Hallen's viewpoint for the potentials along the axis of symmetry.*

Instead of discussing the integral equation (77) or (100), we may revert to (64) and start from the following condition:

$$F_x = \frac{\mu_0}{4\pi}\int I(x')G(r)dx' = \text{sinusoidal function.} \quad (107)$$

What sort of sinusoidal function shall we use? This depends upon the type of problem: receiving antenna or transmitting antenna. Let us assume a transmitting antenna fed at the middle point $\frac{1}{2}l$ (conditions 77b, II-b, §12) with a driving electromotive force V_0. The problem will be more conveniently stated if we take the origin on the middle point of the antenna, which will now extend from

$$-h = -\tfrac{1}{2}l \quad \text{to} \quad h = \tfrac{1}{2}l.$$

The sinusoidal potential distribution corresponding to the transmitting antenna is

$$F_x = -\frac{i}{c}\left(C_1 \cos kx + \frac{1}{2}V_0 \sin k|x|\right). \quad (108)$$

The cosine term yields no discontinuity at $x=0$ but the sine term is discontinuous on $\partial F_x/\partial x$

$$\frac{\partial F_x}{\partial x_{+0}} - \frac{\partial F_x}{\partial x_{-0}} = -i\frac{k}{c}V_0 \quad (108a)$$

which, according to (106), corresponds to the discontinuity V_0 in the distribution of the scalar potential V. This gives exactly what we need, namely, a driving electromotive force V_0 at $x=0$. So, according to Hallen and R. King, the fundamental equations involved are (107) and (108), and the problem is to find the corresponding current distribution $I(x)$. The limits of the integral (107) are now $-h$ and $+h$ instead of 0 and l.

In this integral, the most important contribution is obtained from the small values of $r = |x - x'|$. Splitting the integral (see eq. 96A):

$$\int_{-h}^{h} I(x')G(r)dx' = \int_{-h}^{h} I(x')\frac{e^{-ikr'}}{r'}dx'$$

$$= I(x)\int_{-h}^{h}\frac{dx'}{r'} + \int_{-h}^{h}\frac{I(x')e^{-ikr'} - I(x)}{r'}dx', \quad (109)$$

$$r' = \sqrt{(x-x')^2 + a^2}.$$

* For a similar result in Schelkunoff's theory see section 17 following eq. (128).

The first integral on the right can be computed immediately

$$\int_{-h}^{h}\frac{dx'}{r'}=\log\frac{\sqrt{(h-x)^2+a^2}+h-x}{\sqrt{(h+x)^2+a^2}-h-x}$$
$$=\Omega+\log\left(1-\frac{x^2}{h^2}\right)+\delta, \quad (110)$$

where

$$\Omega=2\log\frac{2h}{a}=2\log\frac{l}{a},$$
$$\delta=\log\left\{\frac{1}{4}\left[\sqrt{1+\left(\frac{a}{h-x}\right)^2}+1\right]\times\left[\sqrt{1+\left(\frac{a}{h+x}\right)^2}+1\right]\right\}.$$

δ is infinitesimal except near both ends of the antenna, where it remains finite.

Substituting (109) and (110) in the original equations (107) (108):

$$I_x=-i\frac{4\pi}{\Omega R_c}\left\{C_1\cos kx+\frac{V_0}{2}\sin k|x|\right\}$$
$$-\frac{1}{\Omega}\left\{I_x\log\left(1-\frac{x^2}{h^2}\right)+I_x\delta\right.$$
$$\left.+\int_{-h}^{h}\frac{I(x')e^{-ikr'}-I(x)}{r'}dx'\right\}, \quad (111)$$
$$R_c=\mu_0 c=376.7=120\pi \quad \text{ohms.}$$

This (but for the omission of a term taking account of the finite conductivity of the metal) is equation (35) of R. King and C. W. Harrison.[21]

Supplementary Considerations

The proof that the potential distribution must be sinusoidal, and the deduction of eq. (108) for the center-fed antenna are the most striking points in R. King's method. Both results apply to the general problem of the cylindrical antenna, as discussed in Sections 12, 13, and are very important.

The main drawback involved in these results is the use of Oseen or Hallen's approximation for the $G(r)$ function (96), discussed in §14 in connection with the problem of end-surfaces. But a most important question arises due to the fact that current and voltage at the driving section are not measured at the same point.

If we take Oseen's approach (97), we compute a fictitious axial current while we measure F and V on the cylindrical surface, $\rho=a$, of the antenna. According to Hallen (98), we measure the superficial current on the cylinder ($\rho=a$) and the potential difference V along the axis.

In both cases, the assumptions are not consistent. V and I are not taken at the same point, and V/I is not a correct definition of the input impedance. Consequently the whole method remains open to criticism inasmuch as the neighborhood of the feeding section ($x=0$) is a region where the field distribution is rather intricate, with very strong fields just where the driving voltage V_0 is applied (on the radius $\rho=a$, Oseen, or $\rho=0$, Hallen), and decreasing fields and voltages adjacent to this point.

It would not be very difficult to improve the results at this point by the method sketched in section 14 (eq. 99). Taking the viewpoint of Oseen, the surface of the antenna is an exact cylinder with rounded ends, and the current I_x computed by means of eq. (111) is a fictitious axial current. But (eq. 99) shows how to compute the actual superficial current. This real current is practically equal to I_x *but for* (1) both ends of the antenna and (2) near the feeding section. The real current on the feeding section should be compared with the driving potential V_0 in order to obtain the actual input impedance.

As for the behavior of the current on and the shape of the rounded ends of the antenna, this point has been discussed by Oseen [17] (p. 10), who shows that the length d of the rounded top (Fig. 37) is shorter than a, the radius of the wire.

Fig. 39 represents an attempt at sketching the field distribution and is intended merely to visualize the criticisms developed above.

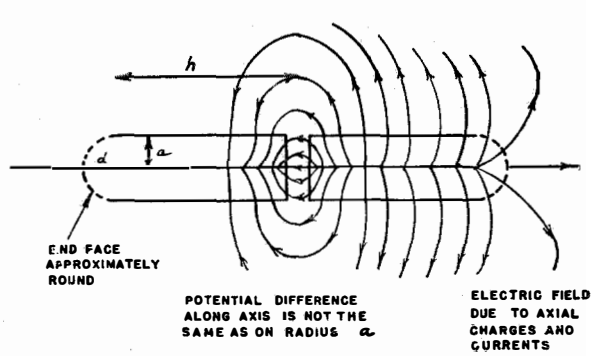

Fig. 39—Approximate field distribution, cylindrical antenna fed in the middle. The longitudinal field on the radius a, in the gap, is compensated by the applied E.M.F.

Let us here introduce the following notations *

$$F_0(x)=\cos kx, \quad G_0(x)=\sin k|x|. \quad (112)$$

The first bracket in (111) will be denoted by I_{x0} and used as a zero order approximation in an expansion in powers of the small quantity $1/\Omega$:

$$I_{x0}=-i\frac{4\pi}{\Omega R_c}\left[C_1F_0(x)+\frac{1}{2}V_0G_0(x)\right],$$

$$I_x=I_{x0}+\frac{1}{\Omega}I_{x1}\cdots$$

$$=-i\frac{4\pi}{\Omega R_c}\left[C_1\left(F_0+\frac{1}{\Omega}F_1\cdots\right)+\frac{1}{2}V_0\left(G_0+\frac{1}{\Omega}G_1\cdots\right)\right]. \quad (113)$$

Substituting in (111):

$$F_1(x)=-F_0(x)\log\left(1-\frac{x^2}{h^2}\right)-F_0(x)\delta-\int_{-h}^{h}\frac{F_0(x')e^{-ikr'}-F_0(x)}{r'}dx' \quad (114)$$

and a similar expression for G_1, when G_0 is substituted for F_0. The C_1 constant is determined by the condition $I_x=0$ for $x=h$.

$$C_1=-\frac{V_0}{2}\frac{G_0(h)+\frac{1}{\Omega}G_1(h)}{F_0(h)+\frac{1}{\Omega}F_1(h)}. \quad (115)$$

Finally

$$I_x=i\frac{2\pi V_0}{\Omega R_c}\frac{\sin k(h-|x|)-\frac{M}{\Omega}}{\cos kh+\frac{A}{\Omega}}, \quad (116)$$

where

$$M=M'+iM''=F_1(x)\sin kh-F_1(h)\sin k|x|+G_1(h)\cos kx-G_1(x)\cos kh,$$

$$A=A'+iA''=F_1(h).$$

This is King and Harrison's formula (39).

The problem is reduced to the computation of the integrals (114) with the F_0 and G_0 functions as point of departure. This is accomplished with a few new approximations, such as neglecting δ, and replacing $(1/r')e^{-ikr'}$ by

$$\frac{1}{|x-x'|}e^{-ik|x-x'|}, \quad r'=\sqrt{(x-x')^2+a^2},$$

* $F_0(x)$ and $G_0(x)$ are R. King's notations, which one should be cautious not to confuse with the vector potential F_z or the $G(r)$ functions used in the previous section. R. King denotes $\frac{1}{2}l$ by h, and k by β.

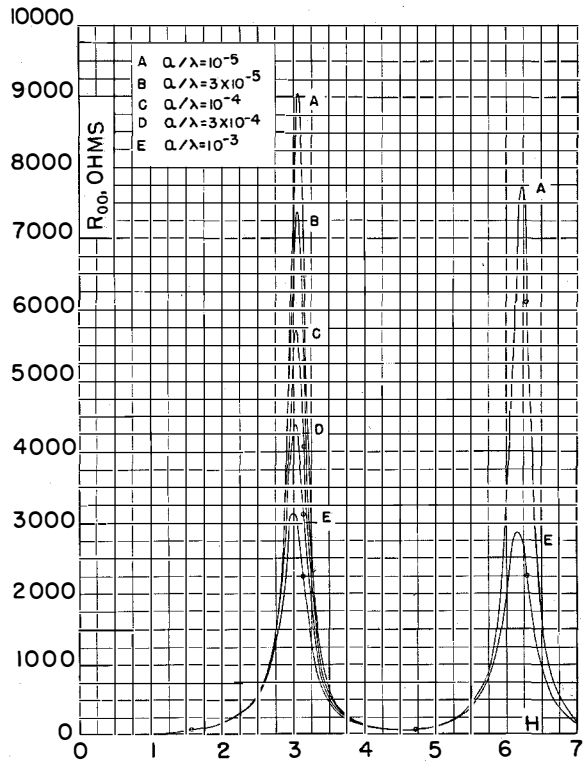

Fig. 40—The input self-resistance R_{00} as a function of $H=2\pi h/\lambda$ with a/λ as parameter. Values of $H=n\pi/2$ are indicated by a small circle.

in a few terms previously grouped in such a way that they practically compensate each other for small values of $|x-x'|$. All these approximations seem perfectly legitimate, and enable one to express the integrals in terms of well-known and tabulated functions, the sine integrals and the cosine integrals. For the detailed computations, the reader is referred to the original papers of Hallen and King.[18, 19]

16. Practical Results for Cylindrical Antennae

All the computations refer to fine antennae where the ratio a/l of radius to length of the cylinder can be neglected. Expansions employ powers of $1/\Omega$ (see §5, eq. 11 or §15, eq. 110);

$$\Omega=2\log\frac{l}{a}.$$

These are the same approximations as in M. Abraham's calculations (or Ryder's) for prolate spheroids.

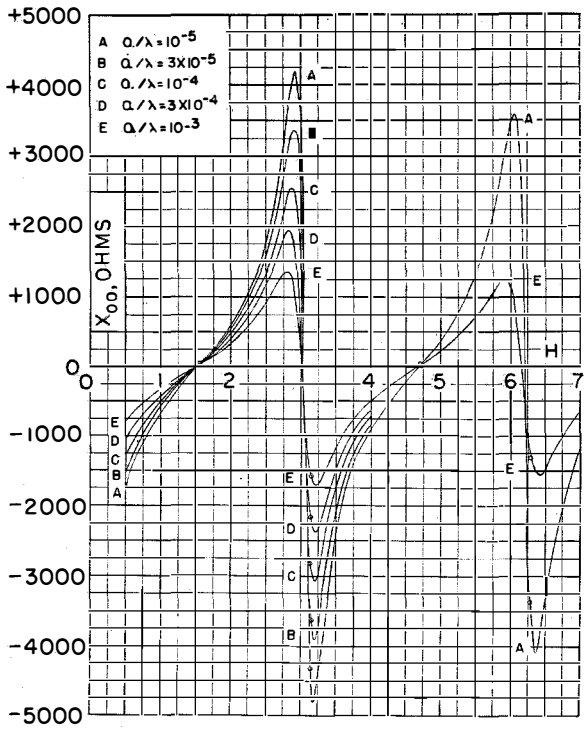

Fig. 41—The input self-reactance X_{00} as a function of $H=2\pi h/\lambda$ with a/λ as parameter. Values at $H=n\pi/2$ are indicated by a small circle.

Free damped oscillations of antenna were first computed by Oseen and Hallen. Oseen obtained only the first approximation, and noticed that the exact shape of the antenna is a cylinder with two rounded end-surfaces whose height d is slightly less than a. Then he finds

$$\lambda_n=\frac{2l}{n}\left[1+\frac{1}{n\pi\Omega}\int_0^{2n\pi}\frac{\sin x}{x}dx\right]$$

proper wave length, (117)

$$\delta_n=\frac{2}{n\Omega}\int_0^{2n\pi}(1-\cos x)\frac{dx}{x}$$

logarithmic decrement,

which can be compared with M. Abraham's results for the ellipsoid: the log decrements are the same in both cases but for ellipsoids λ_n is $2l/n$ with no correction in $1/\Omega$. Hallen performed the same calculations up to the next approximations in $1/\Omega^2$. He proved that the correction in $1/\Omega$ on λ_n drops out for ellipsoids, and checked Abraham's results for this special shape. As for cylindrical antennae, his final formulae read:

$$\lambda_1=\frac{2l}{1}\left[1+\frac{0.4514}{\Omega}+\frac{3.31}{\Omega^2}\cdots\right],$$

$$\lambda_2=\frac{2l}{2}\left[1+\frac{0.2375}{\Omega}+\frac{2.10}{\Omega^2}\cdots\right],$$

$$\lambda_3=\frac{2l}{3}\left[1+\frac{0.1611}{\Omega}+\frac{1.58}{\Omega^2}\cdots\right],$$

$$\lambda_n=\frac{2l}{n}\left[1+\frac{1}{\Omega}\left(\frac{1}{2n}-\frac{1}{2n^2\pi^2}\right)+\frac{1}{n\Omega^2}(1.5\log n+2.93)\right], \quad (118)$$

$$\delta_1=\frac{4.875}{\Omega}+\frac{11.71}{\Omega^2}\cdots,$$

$$\delta_2=\frac{3.114}{\Omega}+\frac{10.38}{\Omega^2}\cdots,$$

$$\delta_3=\frac{2.344}{\Omega}+\frac{9.15}{\Omega^2}\cdots,$$

$$\delta_n=\frac{2\log n+4.83}{\Omega}+\frac{3(\log n)^2+11.72\log n+9.98}{n\Omega^2},$$

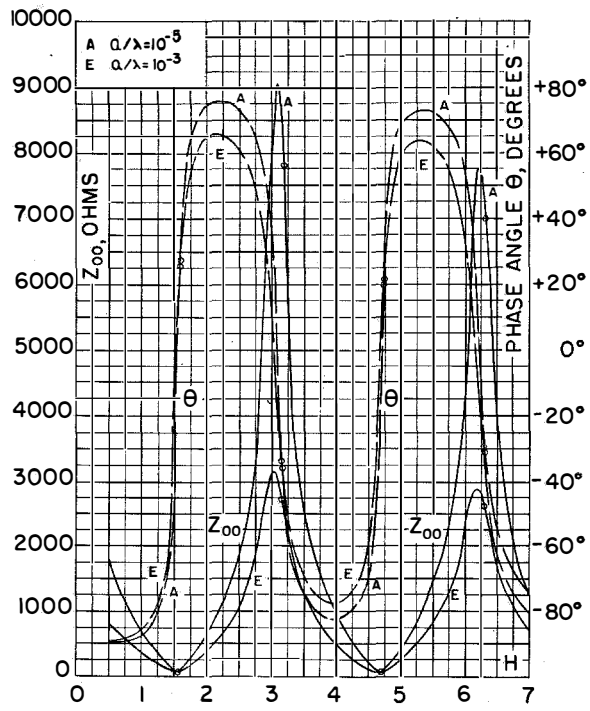

Fig. 42—The magnitude of the input self-impedance $|z_{00}|$ and its phase angle θ as functions of $H=2\pi h/\lambda$ with a/λ as parameter. Values at $H=n\pi/2$ are indicated by a small circle.

where $n>3$. The first approximation in $1/\Omega$ checks with Oseen's formulae (117). The second approximation $(1/\Omega^2)$ is obtained with Hallen's assumptions but it is not certain that Oseen's method would yield exactly the same results. Nevertheless, it is generally agreed that ellipsoidal antennae characteristically require no $1/\Omega$ correction of the λ_n terms.

Transmitting Antennae were very carefully discussed by Ronold King. We shall now summarize his results, keeping in mind the criticism in Section 15 (Fig. 39) which should be of special importance in the neighborhood of the antiresonance frequencies ($l\approx\lambda$, for instance). The theory yields the following formula for the impedance, Z_{00}, across the gap where the antenna is fed:

$$Z_{00}=-i\Omega 60\frac{\cos H+\frac{\alpha_1}{\Omega}+\frac{\alpha_2}{\Omega^2}}{\sin H+\frac{\beta_1}{\Omega}+\frac{\beta_2}{\Omega^2}\cdots}, \qquad (119)$$

$$H=2\pi\frac{h}{\lambda}=\pi\frac{l}{\lambda}$$

where $\alpha_1\alpha_2\beta_1\beta_2$ are functions to be computed from Hallen's theory.

Ω is a quantity ranging from 10 to 30 in most practical antennae (h/a ranging from 10 to 10^8); approximations should hold good for large values of Ω. R. King also took into account the finite conductivity of the metal (copper) and found that it did not play any significant role, so that most results can be computed practically for an ideal metal.

The usual assumption made in the elementary theory is that of sinusoidal current distribution. It can be used only if Ω is so large that all terms in α_1/Ω, α_2/Ω^2, β_1/Ω, β_2/Ω^2 can be neglected in eq. (119).

A better approximation is obtained on the assumption of the following inequalities

$$|\Omega\sin H|\gg\beta_1, \quad |\Omega\cos H|\gg\alpha_1,$$

which yield

$$Z_{00}=-i60\left[\Omega\operatorname{cotg}H+\frac{\alpha_1}{\sin H}-\beta_1\frac{\cos H}{\sin^2 H}\cdots\right]$$
$$=R_{00}+iX_{00}. \qquad (120)$$

This approximation leads to a radiation resistance R_{00} which is practically independent of the radius, a result which we already found to apply for fine antennae. It does not hold for thicker antennae, where higher terms should be kept in the expansion of Z_{00}. All data computed by R. King and F. G. Blake [20] were based on the assumption that terms in α_1/Ω and β_1/Ω should be kept in eq. 119, but that α_2/Ω^2 and β_2/Ω^2 could be dropped. They did not rely on the simplified formula (120).

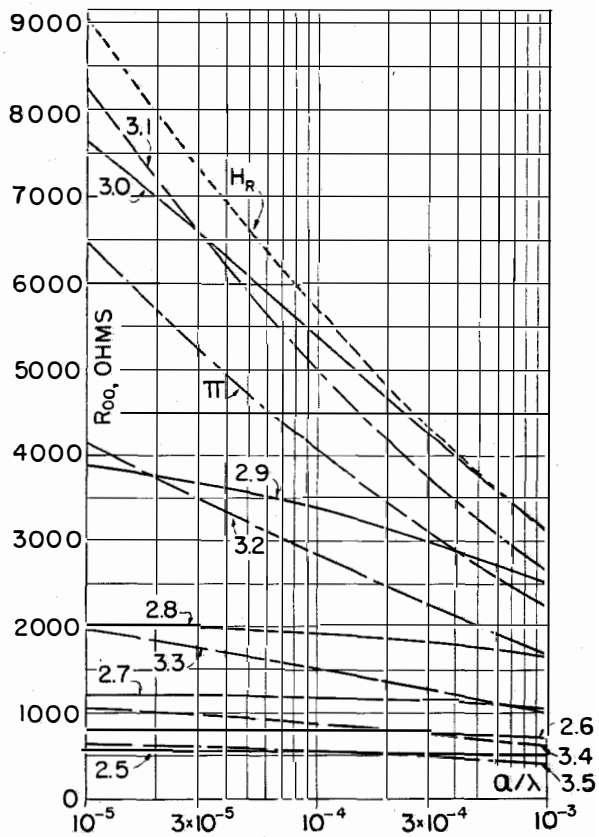

Fig. 43—The input self-resistance R_{00} as a function of a/λ with $H=2\pi h/\lambda$ as a parameter. The curve marked H_R is for values of R_{00} when H is adjusted to antiresonance near $H=\pi$ for indicated values of a/λ.

Curves for R_{00} and X_{00} as functions of H, with a/λ as parameters, are reproduced in Figs. 40 and 41 (λ, wave length; a, radius of wire; $2h$, length of antenna; $H=2\pi h/\lambda$). Fig. 42 is a plot of the magnitude $|Z_{00}|$ of the input self-impedance and of the phase angle θ as functions of H, with a/λ as parameters. Figs. 43, 44, 45 give R_{00} and X_{00} as functions of a/λ with H as a parameter. If a given e.m.f. is applied on the antenna, the maximum current intensity is obtained when the frequency (or λ) is such as to make $|Z_{00}|$ a minimum (see Fig. 42). This does not correspond to a phase angle θ zero. Which

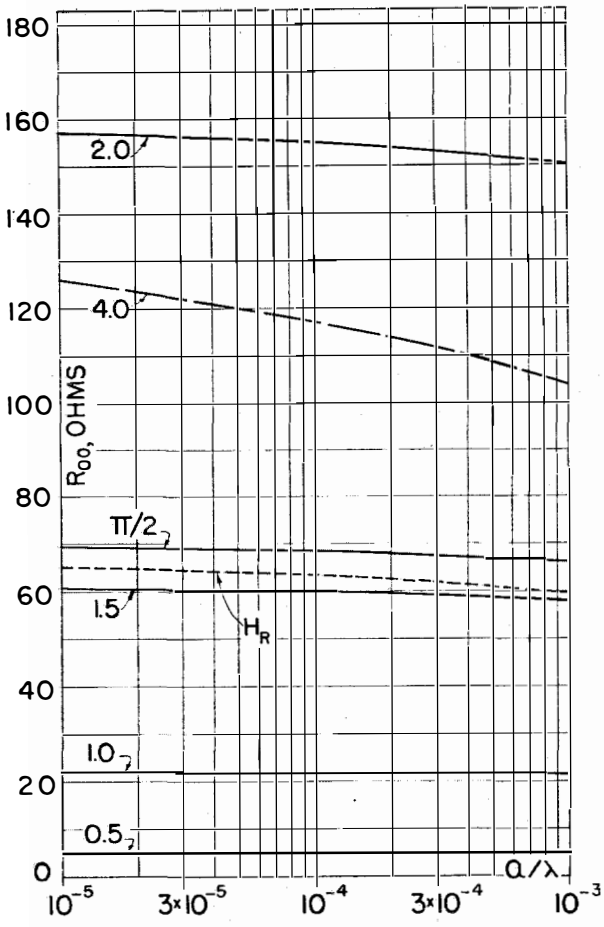

Fig. 44—Same as Fig. 43 for low-resistance ranges. The curve marked H_R is for values of R_{00} when H is adjusted to resonance near $H=\pi/2$ for indicated values of a/λ.

of these two conditions should be chosen to define "resonance"? R. King and Blake use the second condition. In a similar way, "antiresonance" is also defined by $\theta=0$, which does not mean $|Z_{00}|$ maximum.

Resonance and antiresonance fall in the neighborhood of $H=n(\pi/2)$, with n odd or even. The exact conditions are found to be

$$\left.\begin{aligned} H_r &= n\frac{\pi}{2}-\delta, \\ \delta &= \frac{Si(2n\pi)}{4\log\frac{n\lambda}{a}}, \qquad n \text{ odd, resonance} \\ \delta &= \frac{4Si(n\pi)-Si(2n\pi)}{4\log\frac{n\lambda}{a}}, \quad n \text{ even, antiresonance} \end{aligned}\right\} \quad (121)$$

where Si signifies the sine integral function.

These δ corrections are plotted in Fig. 46 as functions of Ω, and the corresponding input resistances at resonance ($n=1$, $n=3$) are shown in Fig. 47, while the input resistances at antiresonance ($n=2$, $n=4$) are found in Fig. 48.

It is clear from Fig. 47 that the usual value of 73.13 ohms taken for the radiation resistance of a symmetrical half-wave antenna is considerably too high. Since the practical range of antennae lies between $\Omega=10$ and $\Omega=30$, the actual values lie between 58 and 68 ohms, depending on the value of a/λ.

The resistance at antiresonance is also found in Fig. 43 (dashed line H_R) and the resistance at resonance in Fig. 44 (dashed line H_R).

In the papers of R. King and Blake [20] and R. King and Harrison,[21] the reader can find many theoretical discussions and numerical data of great importance. The second paper contains graphs of the different functions involved in the calculations, and curves representing the current

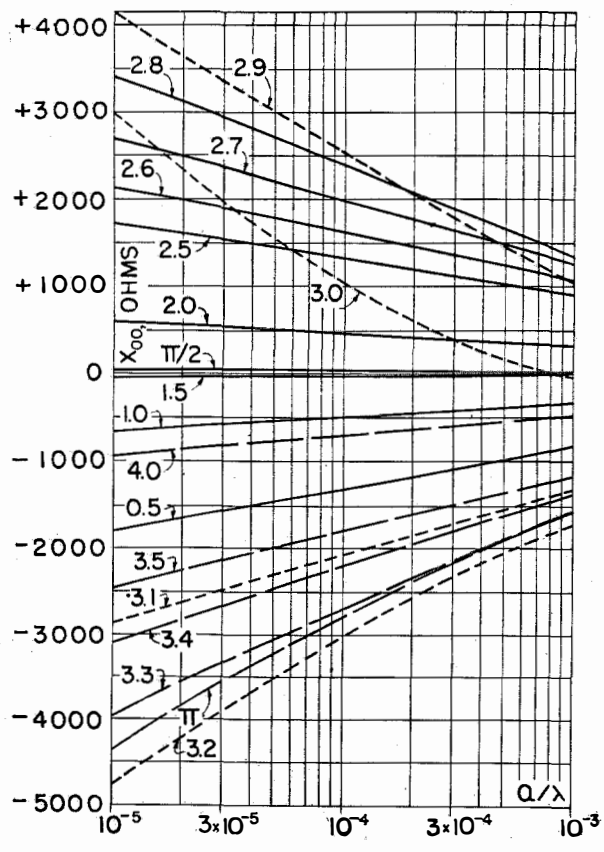

Fig. 45—The input self-reactance X_{00} as a function of a/λ with $H=2\pi h/\lambda$ as parameter.

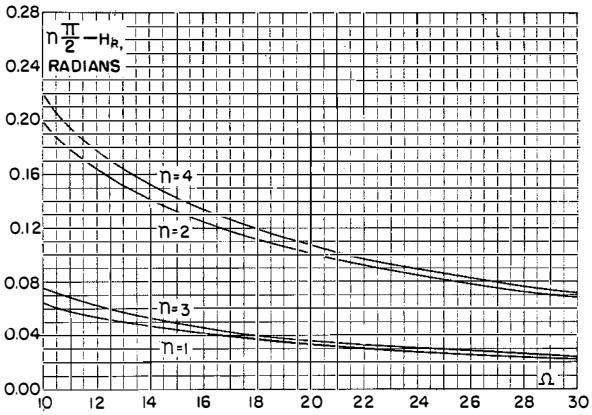

Fig. 46—Resonant lengths H in radians for zero radius minus resonant lengths in radians for non-vanishing radius, viz., $n\pi/2-H_r=2\pi(n/4-H_r/\lambda)$, as functions of $\Omega=2\log(2h/a)$ for $n=1, 2, 3, 4$.

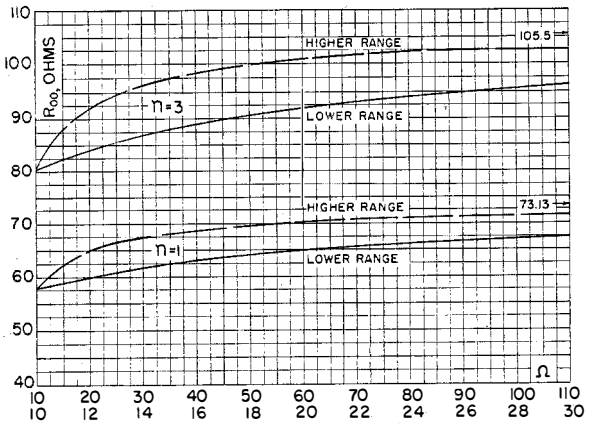

Fig. 47—The input self-resistance R_{00} at the $n=1$ and $n=3$ resonant lengths as a function of $\Omega=2\log(2h/a)$.

distribution along the antenna, in order to show how much it differs from the sinusoidal curves. (The present author is greatly indebted to Prof. R. King for furnishing a copy of this second paper prior to publication.)

The theoretical results of R. King do not seem to check very well with experimental data, especially near the frequencies of antiresonance where the resistance obtained is still too high. (For $l/a=60000$, the maximum resistance is 8400 ohms as compared with the experimental 6150 ohms, as pointed out by Schelkunoff.) This discrepancy would seem to be explained by the discussion contained in the preceding section.

17. The Biconical Antenna

A different method, for the discussion of the theory of antennae, has been adopted by Barrow, Schelkunoff and other writers.[22,23] It is based on the model of a biconical structure fed at the common apex. This structure is very well adapted to the theoretical discussion, especially in the case of fine antennae since it makes possible comparison with lines or cables terminated on a certain given impedance. Once this biconical antenna has been analyzed and all its properties understood, the results can be extended (at least approximately) to more complicated structures and shapes. This biconical model was very carefully studied in a paper by Schelkunoff [23] and in the book [24] recently published by the same author. We shall refer to this paper and this book for all details of mathematical calculations and concentrate on an analysis of the fundamental assumptions, in order to note the exact role played by certain simplifications and approximations, and to show

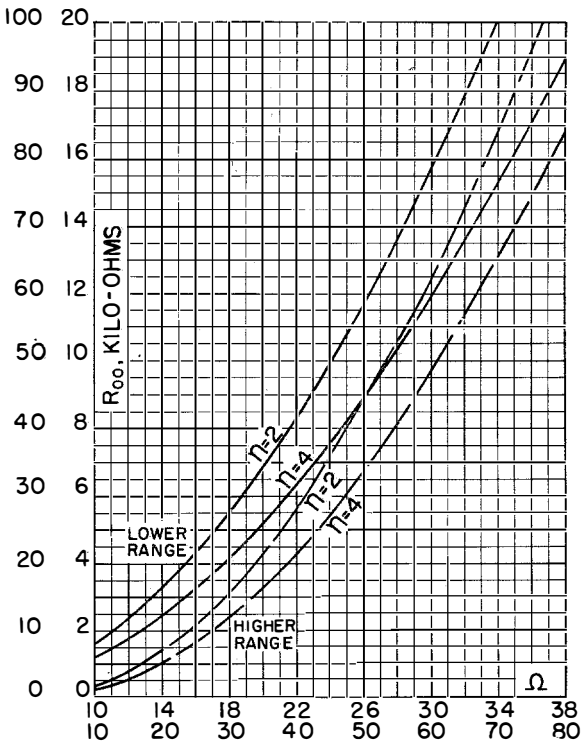

Fig. 48—The input self-resistance R_{00} at the $n=2$ and $n=4$ antiresonant lengths as functions of $\Omega=2\log(2h/a)$.

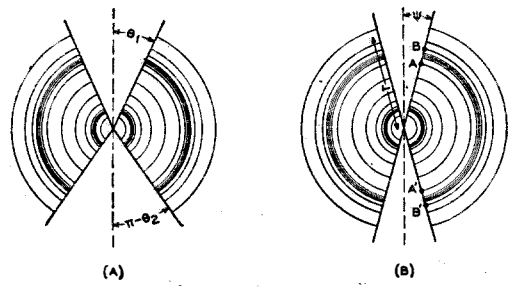

Fig. 49—Cross sections of infinitely long conical conductors and electric lines of force for principal waves.

the connection between this method and the ones discussed in the preceding sections. This will also reveal the type of corrections or amplifications that might be needed.

Let us first state the characteristic feature of biconical antennae when compared to conventional lines: antennae are essentially *multiple* transmission lines and *not simple* like the single mode transmission line. Hence, antennae should be compared to wave-guides with their many different modes of wave transmission. The comparison also can include coaxial cable at ultra-high frequencies when it becomes necessary to take into account, in addition to the fundamental type of wave, the higher wave types similar to those found in wave-guides (Stratton,[7] p. 547).

The principal wave of a biconical device can be found very simply. Let us assume a structure like the one in Fig. 49 with two coaxial conical conductors (perfect metal, infinite conductivity) energized at the common apex. If the cones were of infinite length the wave should be such that the lines of electric force would follow the meridians of spheres concentric with the apex of the cones. This is the principal wave, of the transverse electric and magnetic type, and it can propagate either outward or inward. The biconical line turns out to be a *uniform* transmission line, with series inductance L and shunt capacitance C per unit length,

$$L=\frac{\mu}{\pi}\log\operatorname{cotg}\frac{\psi}{2},\quad C=\frac{\pi\epsilon}{\log\operatorname{cotg}\frac{\psi}{2}},\qquad(122)$$

$$K=\frac{\eta}{\pi}\log\operatorname{cotg}\frac{\psi}{2}\approx 120\log\operatorname{cotg}\frac{\psi}{2},$$

when ψ is the common angle of both cones (Fig. 49B); K is the characteristic or surge impedance of the line; and $\eta=\sqrt{\mu/\epsilon}\approx 120\pi$, ϵ, μ being taken in rational Giorgi units[24] (p. 287). For this principal wave, currents and voltages are related by the usual formulae:

$$\left.\begin{aligned} V_0(r)&=V_{0+}e^{i\omega t-ikr}+V_{0-}e^{i\omega t+ikr},\quad k=\frac{\omega}{c}=\frac{2\pi}{\lambda}\\ I_0(r)&=I_{0+}e^{i\omega t-ikr}+I_{0-}e^{i\omega t+ikr},\\ V_{0+}&=KI_{0+},\quad V_{0-}=-KI_{0-},\end{aligned}\right\}(123)$$

where I_{0+}, V_{0+} represent a wave spreading from the apex and I_{0-}, V_{0-}, a reflected wave propagating inward. The voltage between the corresponding points AA' is defined as the line integral of the electric field along any line lying completely in the sphere passing through AA'.

For a small angle ψ,

$$K=120\log\frac{2}{\psi}=120\log\frac{2r}{\rho}\left\{\begin{array}{l}\frac{r}{\rho}=100\\ K=635\text{ ohms}\\ \frac{r}{\rho}=1000\\ K=913\text{ ohms}\end{array}\right.\qquad(124)$$

where ρ is the radius of the cone, measured at the distance r.

Instead of infinite cones, let us now assume a reflecting sphere at $r=l$, with some finite impedance Z_t; standing waves then will result from the superposition of waves travelling in both directions, as in (123).

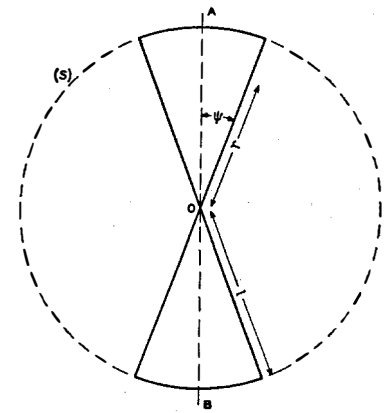

Fig. 50—The cross section of a conical antenna of length l and of the "boundary sphere" S.

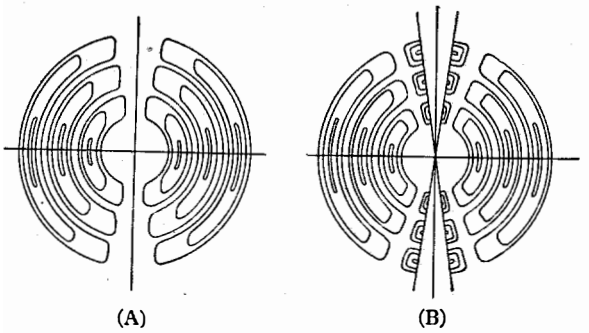

Fig. 51—Electric lines for the first-order transverse magnetic spherical waves: (A) lines in free space (B) lines in the presence of two coaxial conical conductors.

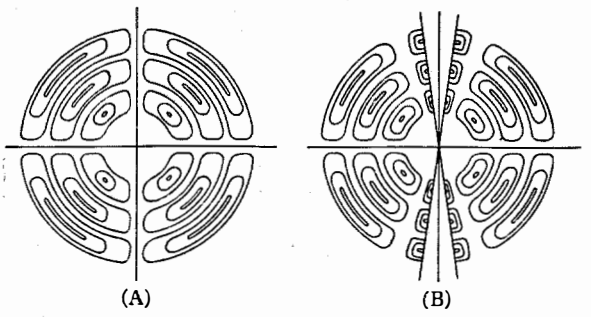

Fig. 52—Electric lines for the second-order transverse magnetic spherical waves: (A) lines in free space (B) lines in the presence of two coaxial conical conductors.

$$
\begin{aligned}
&I_0(r) = I_{0s}\sin k(l-r) + I_{0c}\cos k(l-r)\\
&V_0(r) = V_{0s}\cos k(l-r) + V_{0c}\sin k(l-r)\\
&\qquad \text{with } e^{i\omega t} \text{ as a common factor}\\
&V_{0s} = -iKI_{0s} \quad V_{0c} = iKI_{0c}\\
&Z_t = \frac{V_0(l)}{I_0(l)} = -iK\frac{I_{0s}}{I_{0c}}\\
&\qquad \text{output terminal impedance,}\\
&Z_i = \frac{V_0(0)}{I_0(0)} = K\frac{-iI_{0s}\cos kl + iI_{0c}\sin kl}{I_{0s}\sin kl + I_{0c}\cos kl}\\
&\text{input impedance at the common apex.}
\end{aligned}
\qquad (125)
$$

These equations would represent the actual conditions in a conical antenna (Fig. 50) if it were not for the fact that the sphere outside the boundary sphere S is a multiple transmission system, with all sorts of transmission modes (spherical waves) which differ from those found in the antenna region. In free space, there is no transmission mode really similar to the principal mode just discussed for the antenna and characterized by its very high field concentrations near the conical conductors. Hence, the energy carried (in the conical antenna region) by the principal wave, from the center to the boundary sphere S, subsequently must travel in different spherical transmission waves. In order to match the field across the sphere S, secondary waves of higher types will also be needed in the antenna region. Schelkunoff's method is essentially based on this matching condition on the surface S. Accordingly, he initially disregards the boundary conditions on both metallic end surfaces, and this is one of the approximations which we shall consider hereafter.

Let us first follow Schelkunoff's discussion and review the different types of spherical waves in free space around the surface S. The principal mode in free space is characterized by electric lines of force whose shape is indicated in Fig. 51A. This is the type of wave emitted from a small doublet at the origin. The second transmission mode is shown in Fig. 52A and corresponds to the wave emitted by a quadruplet. This wave will not occur in our problem for reasons of symmetry, but the next one will be excited. Mathematically these different modes are represented by zonal harmonics, the radial electric field being proportional to $P_n(\cos\theta)$ and the meridian field to $dP_n/d\theta$, where P_n is the nth Legendre polynomial. The symmetry of the antenna problem brings out the odd harmonics only ($n=1, 3, 5 \cdots$).

Similar higher modes can be found in the antenna region (between both metallic cones) and are shown in Fig. 51B and 52B. An important result is pointed out by Schelkunoff, namely: For all secondary waves, in the biconical antenna region, the voltage between two points AA' on the upper and the lower cone is zero:

$$V_m(r) = 0. \qquad (126)$$

This voltage is defined, as in the previous case of the principal wave, as the line integral of the electric field along a path joining AA' and entirely situated in the equiphase surface. Further, for any secondary wave, the conduction current vanishes at the origin:

$$I_m(0) = 0. \qquad (127)$$

(This current actually varies as $r^{m+1+120/K}$ near the origin.)

Both results have a very simple physical meaning and can readily be explained. Referring to Fig. 51B or 52B, imagine arrows marking the direction of the electric field along each closed line of the figure. If we follow a circle AA' running from the upper to the lower cone, we cross as many regions with positive fields as with negative fields, and these different intervals cancel each other. On the contrary, in the case of the principal V_0 wave, the electric field always points in the same direction along the AA' circle (Fig. 50) and V_0 does not vanish. This explains the general result (126).

As for (127), it means that higher modes of transmission correspond to quadripoles, sextupoles and all multiplets of higher order at the origin; and it is well known that, for such higher multiplets, the total current is zero, the doublet being the only structure to yield a finite current at the origin.

We can now write the complete voltage-current equations for the biconical antenna energized at the center:

$$\left.\begin{aligned} V &= V_0(r) \\ I(r) &= I_0(r)+\hat{I}_1(r)=I_0(r)+I_1(r)+I_3\cdots \\ \hat{I}_1(0) &= 0 \end{aligned}\right\} \quad (128)$$

where $V_0 I_0$ are given by (125). The total voltage wave consists of just two principal waves, showing incomplete reflection at the sphere S; the current distribution is much more elaborate.

This first result checks very exactly with the statement of R. King (see section 15, eq. 105) that the voltage distribution along a fine antenna

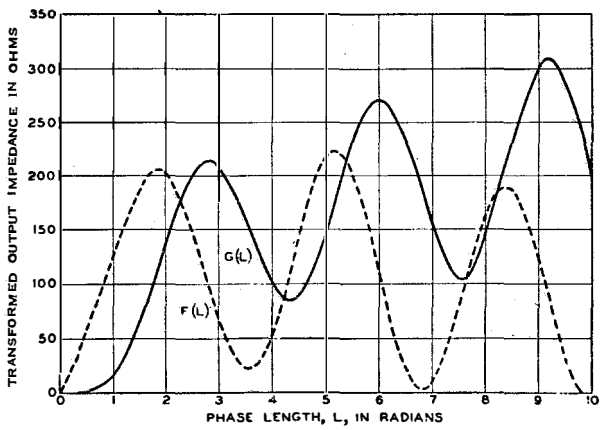

Fig. 53—The real and the imaginary parts of the "transformed" output impedance $K^2/Z_t = G(L)+iF(L)$, where $L=2\pi l/\lambda$.

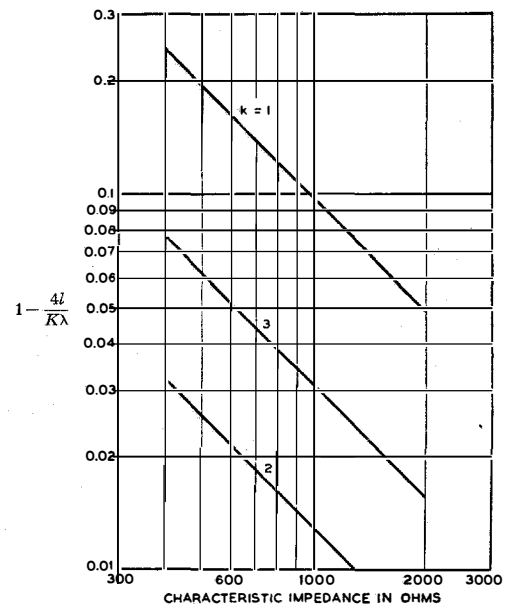

Fig. 54—Deviation of the resonant length of conical antennae from $2l=k\lambda/2$.

must always be sinusoidal, while the current distribution may appear as much more complicated. Both authors agree completely on this important point.

The next point to be emphasized is that secondary waves in an antenna affect the amplitudes of the two principal waves in the same way as an output impedance.

Consider the principal waves in the antenna as given by (125). Substituting from (128) in the expression for the output impedance, Z_t,

$$Y_t=\frac{1}{Z_t}=\frac{I_0(l)}{V_0(l)}=\frac{I(l)}{V(l)}-\frac{\hat{I}(l)}{V(l)} \quad (129)$$

which suggests a termination of the line with two shunt admittances I/V and $-\hat{I}/V$ in parallel. The total current $I(l)$ may be and generally is different from zero. For instance, if the tops of the conical conductors are large, an appreciable current may flow over the edge in order to load or discharge the capacity represented by both end-surfaces. However, if the cross-section of these end-faces is very small, the total current $I(l)$ is zero and

$$I_0(l)=-\hat{I}(l), \quad Y_t=\frac{1}{Z_t}=-\frac{\hat{I}(l)}{V(l)}. \quad (130)$$

This means,* as in the case of the cylindrical antenna previously discussed, the neglect of small terms in ρ/l (ρ, radius of the cones at the top) as compared with logarithmic terms represented by the K surge impedance (124), which plays a similar role to the quantity Ω (eq. 11).

Assuming zero current on top of the cones (130) and fine antennae (large K), and matching the fields over the open part of the sphere S:

$$\frac{I_{0c}}{I_{0s}}=\frac{F(L)-iG(L)}{K}, \tag{131}$$

where L is the "phase length" of each cone

$$L=2\pi\frac{l}{\lambda} \tag{132}$$

while F and G are derived by series expansions in terms of Bessel functions. Substituting in (125) for the *output impedance*,

$$Z_t=\frac{K^2}{G+iF}, \quad \frac{K^2}{Z_t}=G+iF. \tag{133}$$

The curves plotted in Fig. 53 show the behavior of these F and G functions with L variable.

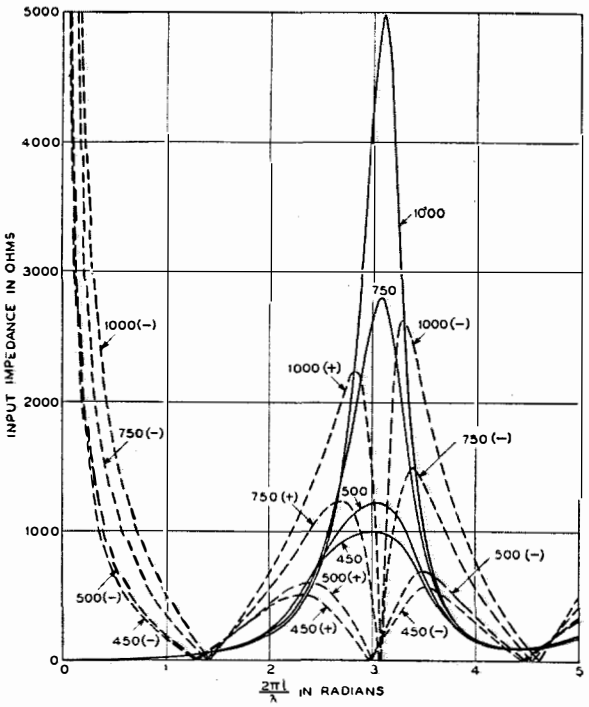

Fig. 55—The input impedance of conical antennae as a function of $2\pi l/\lambda$ and K. Solid curves represent the real component and the dotted curves the imaginary.

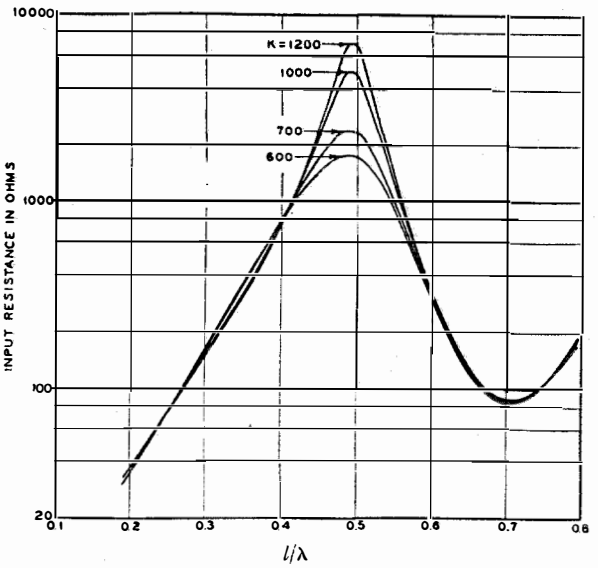

Fig. 56—The input resistance of conical antennae.

The following expressions represent a fairly good approximation for fine conical antennae:

$$G=60[C+\log 2L-Ci2L]+30[C+\log L-2Ci2L+Ci4L]\cos 2L+30[Si4L-2Si2L]\sin 2L\cdots, \tag{134}$$

$$F=60Si2L+30[Ci4L-\log L-C]\sin 2L-30Si4L\cos 2L\cdots,$$

where $C=0.577$ is the Euler-Mascheroni constant, and Ci and Si the cosine and sine integrals (shapes other than biconical would especially modify the F function; G is rather insensitive to such changes).

From the terminal or output impedance Z_t, the *input impedance* as observed between both apexes of the cones can now be computed:

$$Z_i=K\times\frac{(G+iF)\cos\left(L-\frac{\pi}{2}\right)+iK\sin\left(L-\frac{\pi}{2}\right)}{K\cos\left(L-\frac{\pi}{2}\right)+i(G+iF)\sin\left(L-\frac{\pi}{2}\right)}. \tag{135}$$

This formula summarizes the whole theory of the biconical antenna.

* This is the viewpoint taken by Schelkunoff in his paper.[23] Later on, in his book[24] Schelkunoff suggests that the same assumption should apply to "hollow conical antennae," but this is not quite justified. In case hollow cones should be considered, it would be necessary to compute the field distribution inside the cone, and a small current $I(l)$ would be found creeping over the edge, requiring special attention.

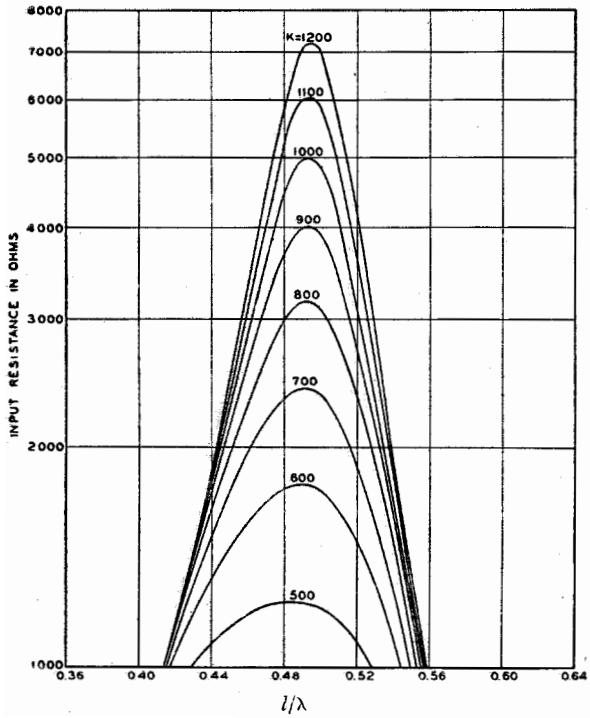

Fig. 57—The input resistance of conical antennae in the neighborhood of the second resonance.

The input reactance (imaginary part of Z_i) vanishes when

$$\tan 2L = -\frac{2KF}{K^2-G^2-F^2} \approx -\frac{2F}{K}\left(\frac{1}{K}\text{ small}\right) \qquad (136)$$

a condition which may be chosen for defining *resonance*. This yields

$$4\pi\frac{l}{\lambda} = 2L \approx n\pi - \frac{2}{K}F\left(\frac{n\pi}{2}\right)\cdots \qquad (137)$$

and shows the influence of the end effect which makes antennae resonate when they are somewhat shorter than $n(\lambda/2)$ (Fig. 54).

The variation of the input impedance of the biconical antenna, as a function of $L=2\pi l/\lambda$, is shown in Fig. 55. Curves 56, 57 and 58 give additional details of the behavior * of input resistance or reactance. It should be noted that the input resistance values obtained in the neighborhood of the second resonance ($l \approx \lambda/2$) are materially smaller than the ones computed by R. King; the order of magnitude predicted by Schelkunoff's theory seems to agree fairly well with experimental data while R. King's values are consistently higher.

18. Comparative Results—Biconical and Straight Antennae

R. King's computations on straight cylindrical antennae embody two weak points, as emphasized in sections 14–16.

I.—No provision is made for the role played by both end surfaces of the cylindrical wire;

II.—Utilization of an approximative G function results in the fact that voltage and current are not computed at the same place, which certainly means an error in the input impedance (see "Supplementary Considerations," section 15, p. 26).

In Schelkunoff's consideration of the biconical antenna, difficulty I relative to the role of end-surfaces still remains, though not exactly in the same manner as in R. King's theory. Schelkunoff's method has the great advantage of avoiding any difficulty in defining input impedance. The biconical shape is very convenient in that it insures an exact definition of voltage and current measurements, both taken exactly at the same place, at the common apex of both cones. Obviously, in actual practice, no antenna is built as a double cone, but this deviation

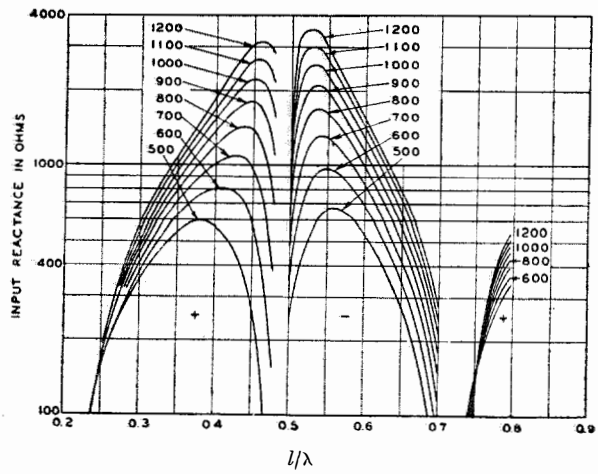

Fig. 58—The input reactance of conical antennae.

* A number of other interesting curves may be found in Schelkunoff's paper, showing the behavior of some important functions, and also visualizing the current distribution along the antenna under different conditions.

seems of minor importance while it certainly is essential in avoiding difficulty II in R. King's method.

To extend the theory, two factors require consideration. Schelkunoff discusses these in his paper and his book, though perhaps rather approximately. A more detailed theory would be helpful.

The first generalization refers to the extension of the theory to non-conical shapes, which are treated as non-uniform transmission lines. The second point is the question of the end-surfaces, the exact role of which can be neglected only for very fine antennae, but must absolutely be taken into account for thick antennae (large ψ angles). A third extension of the theory may be found in Schelkunoff's work—the case of two cones of different lengths.

Let us first summarize the discussion of antennae of various shapes. They are assumed to be sufficiently fine, and each small segment of the antenna is regarded as a section of a cone, with the following approximate expression for the distributed series inductance and shunt capacitance per unit length (cf. eq. 122–124):

$$L=\frac{\mu}{\pi}\log\frac{2}{\psi}=\frac{\mu}{\pi}\log\frac{2r}{\rho},\quad C=\frac{\pi\epsilon}{\log\frac{2r}{\rho}}, \tag{138}$$

where ρ = radius of the antenna at the distance r. Hence, the transmission equations for the *principal waves* can be written:

$$\frac{dV}{dr}=-i\omega LI,\quad \frac{dI}{dr}=-i\omega CV,$$
$$\frac{d}{dr}\left(\frac{1}{C}\frac{dI}{dr}\right)=-\omega^2 LI, \tag{139}$$

which as a first approximation yield

$$V(r)=\sqrt{K}(Ae^{-ikr}+Be^{ikr}),$$
$$I(r)=\frac{1}{\sqrt{K}}(Ae^{-ikr}-Be^{ikr}), \tag{140}$$

with

$$K(r\cdot\rho)=\frac{1}{\pi}\sqrt{\frac{\mu}{\epsilon}}\log\frac{2r}{\rho}\approx 120\log\frac{2r}{\rho}\quad\text{ohms.}$$

Starting from these expressions, one may achieve a closer approximation and obtain, for instance, the resonance conditions for fine antennae of various shapes [23] [§IV, p. 506]. The discussion shows that "equivalent antennae" of given length but different cross-sections and shapes can not be derived from geometrical drawings (see section 6, Fig. 14), but necessitate the application of rather intricate formulae involving the different parameters.

The approximations developed in this discussion may seem reasonable at first sight, but one link is missing: no proof is brought forth that the method utilized actually represents a first approximation to the real problem based on Maxwell's equations. Such a proof is most desirable, as it also would furnish a possible means of computing the order of magnitude of the neglected terms, as well as give a basis for possible corrections or amplifications.

A fundamental modification of the point of departure will probably be necessary since equations (140) do not check with the general result obtained by R. King (section 15, eq. 105) that the voltage distribution along a cylindrical antenna must always follow an exact sinusoidal law, while only the current distribution should depart from a sine curve. On the contrary, eq. (140) shares the correction ($\sqrt{K}$ coefficient) between V and I in a symmetrical way, which throws some suspicion on the validity of the whole procedure. Further comparison of eq. (139) may be made with the real equation for wave propagation along a straight wire, as found in sections 12–13. It was shown in section 12 how difficult it was to reduce the wave propagation equation along a wire to the conventional cable equation. Reduction led to eq. (78), which differs materially from (139). Hence the point of departure of Schelkunoff's approximations with the conventional cable equations (139) seems hard to justify, and his whole discussion of antennae of arbitrary cross-section seems highly provisional.

Starting from (140), Schelkunoff defines an average characteristic impedance *

$$K_a=\frac{1}{l}\int_0^l K(r\cdot\rho)dr$$

which, for a straight cylindrical wire, becomes

$$K_a=120\left(\log\frac{2l}{a}-1\right) \tag{141}$$

* One should be very cautious never to use formula (141) in equations containing K^{-1}, K^{-2} or generally K^n, since $\overline{(K^n)}\neq(\bar{K})^n$.

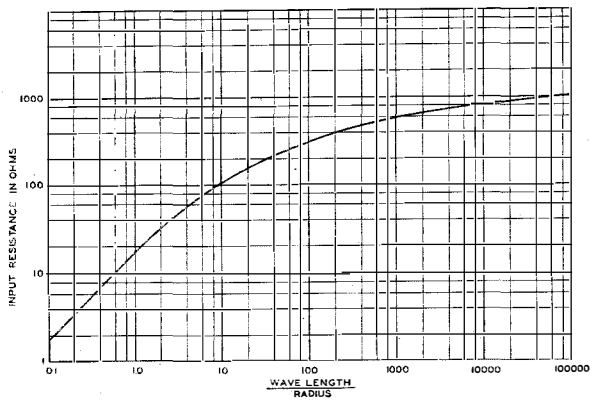

Fig. 59—The input impedance of the infinitely long cylindrical wire. For $0.1<\lambda/2a\leqq 3$, this curve has been computed by numerical integration from (140); for $\lambda/2a \geqq 500$, the curve has been computed from the approximate equation (141); then, the two curves have been freely joined together.

and discusses the properties of cylindrical antennae of various l/a ratios (various K_a values). Fig. 59 represents a graph of eq. (141), Fig. 60 gives the input resistance and Fig. 61 the input reactance of cylindrical antennae in free space for different values of the ratio l/λ. Despite the serious criticisms which can be made of the very foundations of this theory, it nevertheless seems that the practical results agree fairly well with experimental data, and that the curves of the preceding figures represent a good approximation of actual facts.

As for the effect of the *end-surfaces*, it is certainly negligible for very fine antennae and would come into play only for rather thick structures. One way to take this effect into account consists in treating the end-surfaces as a small capacity added in parallel with the terminal impedance which, according to Schelkunoff, represents the radiation. This is the method sketched in section 13 (eq. 85–88) and used by Schelkunoff in his book (p. 465). It shows that the effect of cap capacitance results in a shortening of the resonant lengths of the antenna. It is negligible for very fine antennae, and may amount to a few per cents for $K=600$, $l/a \approx 100$.

More rigorous analysis would necessitate reconstruction of the theory for the whole radiating system, consisting of the two cones and the spherical caps. This represents a very interesting configuration for a broad-band antenna. In section 5, 7 and 8 spherical and spheroidal antennae were discussed, assuming that they were fed along their equatorial circle. For materializing such a system, one of the best approaches consists in the biconical structure, connected at the end of a coaxial line and terminated with large spherical caps. The complete theory can be formulated in a manner similar to Schelkunoff's procedure and summarized above.

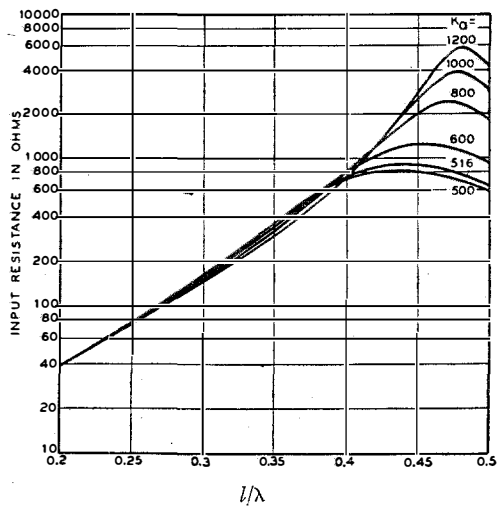

Fig. 60—The input resistance of cylindrical antennae in free space. For vertical antennae over a perfectly conducting ground divide the ordinates and K_a by 2.

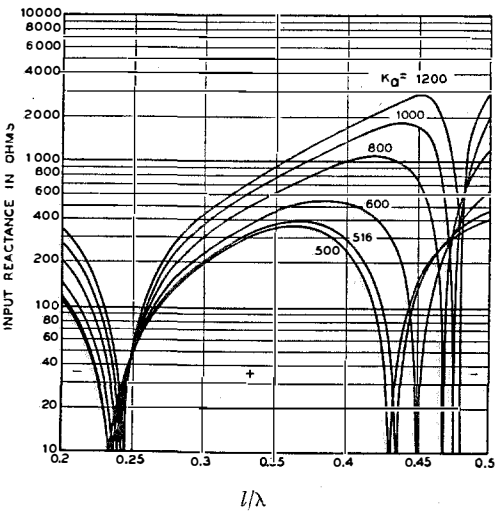

Fig. 61—The input reactance of cylindrical antennae in free space. For vertical antennae over a perfectly conducting ground divide the ordinates K_a by 2

Instead of computing the spherical waves in free space (Fig. 51A or 52A), one should adopt the conception utilized by Stratton and Chu (see section 8) of the spherical waves around a conducting sphere. These waves satisfy the boundary conditions on the spherical caps, and must be joined with the waves found inside the sphere (*S* of Fig. 50) in the biconical region. While such calculations may be found in Schelkunoff's book [24] (p. 472), they have been per formed only approximately, and no curves or tables to aid in visualizing the results are included.

19. Conclusion

The main results of this critical survey of the theory of antennae can be summarized concisely:

A—*Reliable theoretical results* have been obtained by different authors for *spherical* and *spheroidal antennae* fed across the equator (see sections 5 and 6). Reliable also are the computations for fine biconical antennae (sections 17 and 18).

B—*Less reliable theoretical data* are those for fine cylindrical antennae. The computations of R. King should be corrected (see sections 15 and 16), after which they certainly could yield excellent results. Schelkunoff's theory of the cylindrical antenna does not seem very reliable, but the results apparently agree with experimental data (section 18).

Many problems remain *to be studied*, especially of broad band antennae. The most important ones would be:

C—*The Thick Cylindrical Antenna* including the effect of end surfaces.

D—*The Spherical or Spheroidal* antenna fed across a finite gap near the equator, or combined with a biconical feeder. (This last problem has been treated, approximately by Schelkunoff.)

E—*The Directivity* of the preceding types of antenna is not exactly known. Polar diagrams of radiation could be drawn from the theory, but this phase of the problem has not been completed, at least in published papers.

A general remark must be added: Of the antenna shapes discussed fully, not one really possesses sharp edges or sharp points. Despite statements made by authors, the theory actually applies to antennae with rounded ends. The present author would suggest that any shape with sharp edges or sharp points should lead to narrow bands. Hence *broad-band* antennae should not only be *thick antennae* (a well known result) but also *rounded* in their shape, i.e., rather *streamlined*. A rigid proof of this recommendation can hardly be given at present.

NOTE: Fig. 34 is reproduced from "On Radiation from Antennas," by S. A. Schelkunoff and C. B. Feldman, *Proc. I. R. E.*, Vol. 30, Nov. 1942; Figs. 49 to 61 inclusive are reproduced from "Theory of Antennas of Arbitrary Size and Shape," by S. A. Shelkunoff, *Proc. I. R. E.*, Vol. 29, Sept. 1941.

For Bibliography see end of first installment of article in "Electrical Communication," Vol. 21, No. 4, 1944.

The Impulse Response of Electrical Networks*

With Special Reference to the Use of Artificial Lines in Network Design

By M. LEVY

Standard Telephones and Cables, Ltd., London

Summary

Part (2) of the paper presents a study of the relations between the attenuation and phase characteristics of 4-terminal networks and their response to an impulse of infinitely short duration applied to the input terminals. It is shown that a rigid relationship exists between the characteristics of 4-terminal networks and their impulse response. In particular:

(*a*) If the phase-shift produced by the network is proportional to frequency, the impulse response is even;

(*b*) If the phase-shift produced by the network is 90° at all frequencies, the impulse response is odd.

From this theoretical analysis is developed, in Part (3), a method of obtaining any desired impulse response by means of a circuit containing a multiple-section artificial line. The response is obtained by effecting appropriate amplitude reflections at the end of each section, these reflections being produced by means of small variable condensers.

An experimental circuit is described enabling characteristics identical with those of any network to be obtained in a few minutes.

A practical example is given of the design of a low-pass filter with linear phase-shift, as used in television circuits.

The paper concludes with the practical consideration of filters having a phase-shift of 90° at all frequencies. It is shown that with the type of circuit described, practical designs are possible for:

(*a*) High-pass filters whose characteristics are satisfactory up to frequencies as high as 10 to 20 times the cut-off frequency;

(*b*) Band-pass filters, if the band width is not too narrow;

(*c*) Low-pass filters, if the frequencies to be transmitted are not too low.

* Reprinted from *Journal of Institution of Electrical Engineers*, Vol. 90, Part III (Comm. Engg.), No. 12, Dec. 1943.

Contents

(1) *Introduction*

THE study which follows originated in a theory published by the author in a series of articles appearing between 1934 and 1937,[1,2,3,4,6] and is the extension of a paper read before the technical staff of the Laboratories of Le Matériel Téléphonique in 1937. This paper dealt with the application of the theory to television and electrical filters and was summarized in an internal memorandum.[5]

The experimental work was carried out in Paris in 1938 and 1939. The research was suspended at the beginning of the war.

The study has been written from memory, all original documents having been lost at the time of the German invasion. For this reason it has not been possible, with a few exceptions, to give photographs of the completed apparatus, but this gap has been filled as far as possible by sketches showing the original apparatus.

Although the author has just learnt of an article by H. E. Kallmann based on the same ideas, it appears to him that the following study preserves its originality as much through its point of view as through the results obtained, for the author seems to be the first to have designed circuits producing satisfactory characteristics and to have indicated the possibility of making networks and filters producing a phase-shift of 90° at all frequencies.

(2) *Theory*

(2.1) *Definition of Networks in Terms of their Impulse Response.*

(2.1.1) *The Impulse Function $I(t)$.*

Let us consider a passive network with a selectivity function $S(\omega)$ and a phase function $\phi(\omega)$. By definition these two functions are such that if a voltage

$$V_0'(\omega) = V_0 e^{i\phi_0(\omega)}$$

is applied at the input of the network, then the output voltage will be

$$V_1'(\omega) = V_1 e^{i\phi_1(\omega)},$$

where $$S(\omega) = V_1/V_0$$

and $$\phi(\omega) = \phi_1(\omega) - \phi_0(\omega).$$

It follows that the behaviour of the network is completely determined by the knowledge of the functions $S(\omega)$ and $\phi(\omega)$.

Hence, two 4-terminal networks having the same selectivity and phase response give the same transformation to an input signal whatever may be their internal constitution.

The behaviour of a network can also be defined by functions other than $S(\omega)$ and $\phi(\omega)$. Heaviside has shown that this system of functions can be replaced by the "indicial response," which is equivalent to the system of functions $S(\omega)$, $\phi(\omega)$ and vice versa. The behaviour of a 4-terminal network is also completely defined by Heaviside's function.

This consideration has led to important conclusions which are well known. One of its most important applications is the study of the transient response of networks. Another, which became important recently with the development of television technique, is the testing of the behaviour of a network by the indicial response. This response can be obtained at once by applying a unit signal at the input of the network. To observe the response on a cathode-ray tube it is preferable to send this signal periodically, i.e. to send "square" signals. By comparing the shape of the response with the square input signal, one can see whether the behaviour of the network is correct.

Several workers have considered, from time to time, another function deduced from Fourier integrals. But as it seemed to lead to nothing more than Heaviside's function it has been neglected. The author has found, however, that it gives a very simple explanation of a series of physical phenomena,[2,3,4,5,6] and it even permits certain generalizations.[2,5,6] He has used this function in the analysis of aperiodic curves, in the study of the distortion produced in the scanning of sound films, and in the study of the distortion produced by television systems.

It is now proposed to apply this function to the definition of 4-terminal networks, to give its physical interpretation and show the practical conclusions which may be arrived at by these means.

Having two functions $S(\omega)$ and $\phi(\omega)$ of the same variable ω, it is possible to derive a third function $I(t)$ of another variable t, which has a rigid relationship with the functions $S(\omega)$ and

$\phi(\omega)$. Knowing $S(\omega)$ and $\phi(\omega)$ it is possible to deduce $I(t)$ and, conversely, knowing $I(t)$ it is possible to deduce $S(\omega)$ and $\phi(\omega)$. The Fourier integral

$$I(t)=\frac{1}{\pi}\int_0^{\infty} S(\omega)\cos[\omega t+\phi(\omega)]d\omega \qquad (1)$$

gives the function $I(t)$ in terms of $S(\omega)$ and $\phi(\omega)$. Conversely, if

$$A(\omega)=\int_{-\infty}^{+\infty} I(t)\sin \omega t dt \qquad (2)$$

and

$$B(\omega)=\int_{-\infty}^{+\infty} I(t)\cos \omega t dt, \qquad (3)$$

the equations

$$S(\omega)=\sqrt{[A^2(\omega)+B^2(\omega)]} \qquad (4)$$

$$\tan\phi(\omega)=\frac{A(\omega)}{B(\omega)} \qquad (5)$$

give $S(\omega)$ and $\phi(\omega)$ in terms of $I(t)$. Thus the function $I(t)$ completely defines the characteristics of the 4-terminal network.

(2.1.2) *Physical Interpretation of the Function $I(t)$. Impulse Response.*

The function $I(t)$ can be regarded as a signal whose spectrum and phase curve are given by Fourier integrals, i.e. by the functions $S(\omega)$ and $\phi(\omega)$. This leads to a very simple physical interpretation of the function $I(t)$.

For this purpose let us apply at the input of the 4-terminal network an electrical impulse of infinitely short duration ϵ (Fig. 1), and let us imagine for the moment that the network can transmit, without amplitude distortion, signals of any amplitude, and the amplitude of the applied impulse to be infinitely great and equal to $1/\epsilon$. On this hypothesis the area of the impulse is equal to unity.

Now, the spectrum of an impulse of infinitely short duration is uniform and comprises all frequencies, all with the same amplitude. This spectrum can be obtained by replacing the function $I(t)$ by the following impulse equation:

$$\left.\begin{aligned} I_i(t)&=\frac{1}{\epsilon} \quad \text{for} \quad -\frac{\epsilon}{2}<t<+\frac{\epsilon}{2} \\ I_i(t)&=0 \quad \text{for} \quad t<-\frac{\epsilon}{2} \quad \text{or} \quad t>+\frac{\epsilon}{2} \end{aligned}\right\} \qquad (6)$$

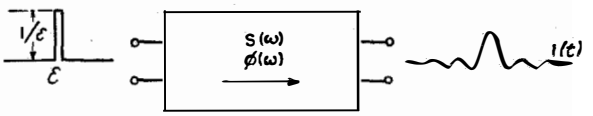

Fig. 1—Showing the close relationship between the selectivity $S(\omega)$ and phase $\phi(\omega)$ functions of a network and its impulse response $I(t)$.

Thus we find:

$$A_i(\omega)=0; \qquad S_i(\omega)=1,$$
$$B_i(\omega)=1/\epsilon\times\epsilon=1; \qquad \phi_i(\omega)=0.$$

Hence, the amplitude of each component of the spectrum is equal to the area of the impulse, i.e. equal to unity for the impulse we have chosen.

To obtain the signal at the output of the network we must multiply each frequency component of the impulse by $S(\omega)$ and add to its phase a phase-shift $\phi(\omega)$. Then

$$S_i(\omega)\times S(\omega)=S(\omega),$$
$$\phi_i(\omega)+\phi(\omega)=\phi(\omega).$$

This gives a signal whose amplitude and phase are given respectively by $S(\omega)$ and $\phi(\omega)$, i.e. by the function $I(t)$, which determines the network. In other words, the function $I(t)$ giving the characteristics of a network is represented by the response of this network to an impulse of finitely short duration and having unit area.

In referring to this property the author will call the function $I(t)$ "the impulse response of the network."

(2.2) *Character of the Impulse Response*

(2.2.1) *Response to a Transient Voltage Function $V(t)$. Transient State.*

Let us apply at the input of the network a voltage $V(t)$ whose amplitude varies with time according to any kind of law, and let us find the response $R(t)$ at the output of the network with the help of the impulse response $I(t)$ of the network.

By definition, if we impress on the network, at the time $t=0$, an impulse $V_{(0)}$ of infinitely short duration dt, we obtain at the output, at time t, a voltage

$$V_{(0)}I(t)dt.$$

Similarly, if we impress at time $t=t_0$ an impulse $V(t_0)$ of infinitely short duration dt, we

obtain at the output, at the mean time t, a voltage

$$V(t_0)I(t-t_0)dt.$$

Adding all the voltages arriving at the output at time t, which are produced by the voltage $V(t)$ impressed on the network, we obtain the response $R(t)$:

$$R(t)=\int_{-\infty}^{+\infty} V(t_0)I(t-t_0)dt_0.$$

To avoid confusion, let us put $t_0=T$; thus, we obtain:

$$R(t)=\int_{-\infty}^{+\infty} V(T)I(t-T)dT. \qquad (7)$$

This equation gives the response of the network for any input voltage as a function of its impulse response.

It leads also to an interesting property of the impulse response. Consider a system of fixed axes and trace the curve $V(T)$ in this system (Fig. 2); consider also a system of axes moving parallel to the preceding ones, the axis of abscissae sliding on the corresponding axis of the preceding system, and trace the curve $I(-t)$ with respect to this new system. Suppose the distance OO_m between the origins of the two systems to be equal to t. If for each value of T we take the product of the ordinates of the two curves, multiply this product by dT and take the sum of these products, we obtain integral (7). Hence to obtain the response of a network at time t, one may superimpose the inverted impulse response $I(-t)$ of the network on the curve representing the input signal, take $OO_m=t$ and integrate as indicated in (7). In moving the inverted impulse response along the input-voltage curve, one may see how the response $R(t)$ of the network varies with time.

From this property we may obtain a simple method for the experimental study of networks in the transient state.

With an impulse generator and a cathode-ray oscillograph we obtain the impulse response of the network. We trace this curve on transparent paper, slide it along a curve representing the input voltage and evaluate integral (7) either qualitatively or quantitatively. This integration can also be made by an electrical device, such as that suggested by the author in an earlier paper.[1, 3]

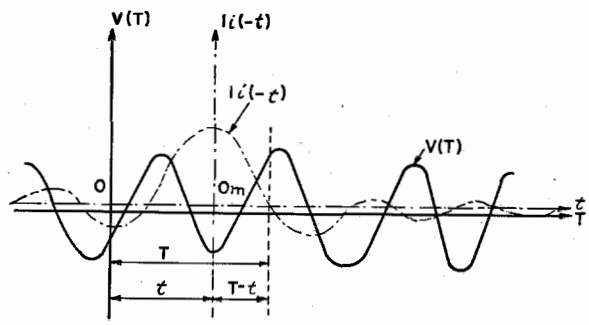

Fig. 2—Graphical method of obtaining the response of a network at time t for an input signal $V(T)$. Superimpose the inverted impulse response $I(-t)$, take $OO_m=t$ and integrate the product $I(-t)V(T)$.

(2.2.2) Relations between the Impulse Response and the Phase Function. Properties of Constant-delay Networks and Networks Producing 90° Phase-shift at All Frequencies.

Between the impulse response $I_i(t)$ and the phase function $\phi(\omega)$ there exists the relation

$$\tan\phi(\omega)=\frac{A(\omega)}{B(\omega)}=\frac{\int_{-\infty}^{+\infty} I_i(t)\sin\omega t dt}{\int_{-\infty}^{+\infty} I_i(t)\cos\omega t dt}. \qquad (5)$$

From this one can draw two important conclusions (*a*) and (*b*):

(*a*) *If the impulse response is even* $[I(t)=I(-t)]$, $A(\omega)=0$ *and* $\phi(\omega)=0$. *The network produces no phase-shift.*

This conclusion can be generalized in the following way. If the impulse response is symmetrical about the axis of abscissae $t=t_0$, the new time variable is $t'=t-t_0$ and equation (1) becomes

$$I(t'+t_0)=\frac{1}{\pi}\int_0^{\infty} S(\omega)\cos[\omega t'+\phi(\omega)+\omega t_0]d\omega$$

and, as $I(t'+t_0)$ is even with respect to the axis $t'=t_0$, we have

$$\tan[\phi(\omega)+\omega t_0]=\frac{\int_{-\infty}^{+\infty} I(t'+t_0)\sin\omega t' dt'}{\int_{-\infty}^{+\infty} I(t'+t_0)\cos\omega t' dt'}=0.$$

Hence $$\phi(\omega)=-\omega t_0.$$

That is, if the impulse response is even with respect to any ordinate, the phase shift produced by the network is either proportional to the frequency or zero.

(*b*) *If the impulse response is odd* $[I(t)=-I(-t)]$, $B(\omega)=0$ *and* $\phi(\omega)=\pm\frac{1}{2}\pi\pm 2m\pi$. *The phase-shift produced by the network is equal to* $\frac{1}{2}\pi(\pm 2m\pi)$ *for all frequencies.*

Here also we can generalize this result in showing that if the impulse response is odd with respect to any ordinate, the phase-shift can be written

$$\phi(\omega)=-\omega t_0\pm\tfrac{1}{2}\pi\pm 2m\pi.$$

In this formula we have added the quantity ωt_0: this expression represents the time which the signal would have taken to pass through the network had it not been deformed by the phase shift $\frac{1}{2}\pi$.

Conditions (*a*) and (*b*) give very simple criteria to determine whether a network produces phase-shifts similar to

$$\phi(\omega)=-\omega t_0\pm 2m\pi \tag{8}$$

or

$$\phi(\omega)=-\omega t_0\pm\frac{\pi}{2}\pm 2m\pi. \tag{9}$$

For this purpose it is sufficient to impress on the network an impulse of very short duration and to observe with a cathode-ray tube whether the response is even or odd.

It will be shown subsequently that this criterion is applicable to the construction of filters having characteristics of the type (8) and (9).

(2.2.3) *Impulse Response of Some Typical Networks. Linear Phase-shift and 90° Out-of-phase Filters.*

Knowing the selectivity function $S(\omega)$ and the phase function $\phi(\omega)$ the impulse response of a network is given by equation (1).

$$I(t)=\frac{1}{\pi}\int_0^\infty S(\omega)\cos[\omega t+\phi(\omega)]d\omega. \tag{1}$$

Integrating the right-hand side for typical networks, we distinguish the following cases: (A) Linear phase-shift networks, and (B) Networks producing a 90° phase-shift.

(*A*) *Linear phase-shift networks.*

For this kind of network

$$\phi(\omega)=-\omega t_0\pm 2m\pi \tag{8}$$

and equation (1) becomes

$$I(t)=\frac{1}{\pi}\int_0^\infty S(\omega)\cos\omega(t-t_0)d\omega. \tag{10}$$

In this formula, substituting for $S(\omega)$ the expressions corresponding to the kind of filter under consideration, we obtain its impulse response:

(*a*) *Ideal low-pass filters.* In this type of filter all frequencies below the cut-off frequency, f_c, are transmitted without amplitude distortion. Hence

$$S(\omega)=1 \quad \text{for} \quad \omega<\omega_c,$$
$$S(\omega)=0 \quad \text{for} \quad \omega>\omega_c.$$

Substitution of this expression in (10) gives

$$I_L(t)=\frac{1}{\pi}\int_0^{\omega_c}\cos\omega(t-t_0)d\omega$$
$$=\frac{\omega_c}{\pi}\cdot\frac{\sin\omega_c(t-t_0)}{\omega_c(t-t_0)}. \tag{11}$$

This curve is shown graphically in (*a*), Fig. 3.

(*b*) *Ideal band-pass filters.* For this type of filter we have

$$S(\omega)=0 \quad \text{for} \quad \omega<\omega_{c_1},$$
$$S(\omega)=1 \quad \text{for} \quad \omega_{c_1}<\omega<\omega_{c_2},$$
$$S(\omega)=0 \quad \text{for} \quad \omega>\omega_{c_2}.$$

Substitution in (10) gives

$$I_B(t)=\frac{1}{\pi}\int_{\omega_{c_1}}^{\omega_{c_2}}\cos\omega(t-t_0)d\omega$$
$$=\frac{\omega_{c_2}}{\pi}\cdot\frac{\sin\omega_{c_2}(t-t_0)}{\omega_{c_2}(t-t_0)}-\frac{\omega_{c_1}}{\pi}\cdot\frac{\sin\omega_{c_1}(t-t_0)}{\omega_{c_1}(t-t_0)}. \tag{12}$$

In this form it may be seen that the impulse response of the band-pass filter, whose cut-off frequencies are f_{c_2} and f_{c_1} $(f_{c_2}>f_{c_1})$, can be obtained by taking the difference between the impulse responses of low-pass filters having cut-off frequencies f_{c_2} and f_{c_1} respectively.

Formula (12) may also be written

$$I_B(t)=\frac{\Delta\omega}{\pi}\frac{\sin\frac{1}{2}[\Delta\omega(t-t_0)]}{\frac{1}{2}[\Delta\omega(t-t_0)]}\cos\omega_0 t. \tag{13}$$

This function is shown graphically in (*b*), Fig. 3.

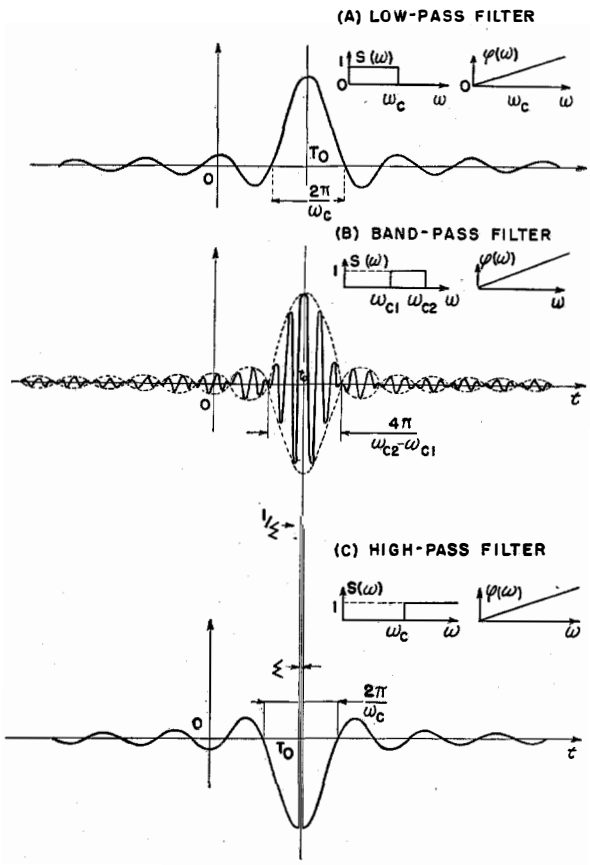

Fig. 3—Impulse response of linear phase-shift filters. These curves are even.

(*c*) *Ideal high-pass filters.* For this type of filter

$$S(\omega)=0 \quad \text{for} \quad \omega<\omega_c,$$

$$S(\omega)=1 \quad \text{for} \quad \omega>\omega_c.$$

In this case the integration of equation (10) presents some mathematical difficulty, due to the fact that the limit of integration is infinity. This is avoided in the following way. If we apply an impulse of infinitely short duration and of unit area to the input of an ideal low-pass filter of cut-off frequency f_c, we obtain at the output a signal whose form is that of the impulse response of this filter. Similarly, if we apply the same signal to the input of an ideal high-pass filter having the same cut-off frequency f_c, we obtain at the output a signal whose shape is that of the impulse response of the filter. It will be clear that the sum of these two signals is a signal whose shape is that of the input signal. If we represent this impulse by the function $I_i(t)$, as in equation (6), we may write

$$I_i(t)=I_L(t)+I_H(t).$$

Hence

$$I_H(t)=I_i(t)-I_L(t). \tag{14}$$

Clearly, this relation holds only when the cut-off frequencies of the two filters are equal.

The curve representing $I_H(t)$ is shown in (*c*), Fig. 3. It will be observed that the amplitude $1/\epsilon$ of the impulse $I_i(t)$ is infinitely great, since ϵ must be infinitely small. But it is clear that, in practice, this condition can only be approximately fulfilled, with consequent detriment to the filter characteristic. We shall return to this question in a later section. Suffice it to say here that this approximation causes a progressive attenuation at high frequencies, with an appreciable effect at frequencies higher than $1/(2\epsilon)$.

(*B*) *Networks producing a 90° phase-shift.*

If $\phi(\omega)=-\omega t_0 \pm 2m\pi \pm \frac{1}{2}\pi$, equation (1) becomes

$$I'(t)=\mp\frac{1}{\pi}\int_0^\infty S(\omega)\sin\omega(t-t_0)d\omega. \tag{15}$$

Substituting for $S(\omega)$, in this equation, the expression appropriate to the type of filter chosen, we obtain its impulse response—just as in the case of linear phase-shift filters. However, the difference between equations (15) and (10) gives rise to important differences in calculation.

In the following we shall only consider ideal filters, i.e. filters having sharp cut-off, with transmission without gain or loss in the transmitting bands and infinite attenuation in the attenuated bands.

(*a*) *Low-pass filters.* Integration of formula (15) gives

$$I_L'(t)=\mp\frac{\omega_c}{\pi}\cdot\frac{1-\cos\omega_c(t-t_0)}{\omega_c(t-t_0)}, \tag{16}$$

$\omega_c/(2\pi)$ being the cut-off frequency of the ideal filter.

This curve is shown in (*a*), Fig. 4. It is useful to note that for values of $\omega_c(t-t_0)$ which are multiples of $2(m+1)\pi$, the curve is a tangent to the envelope

$$E(t)=\mp\frac{2}{\pi(t-t_0)}. \tag{17}$$

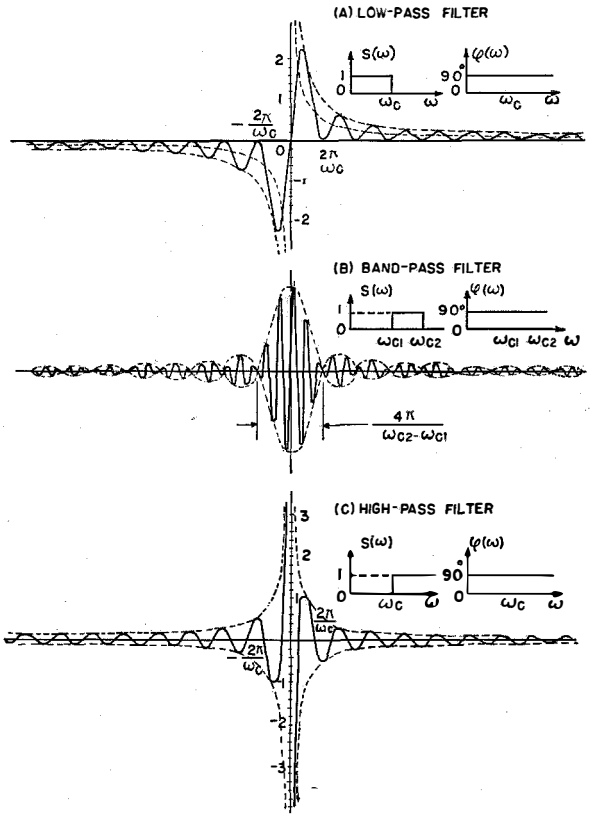

Fig. 4—Impulse response of filters producing 90° phase-shift at all frequencies. These curves are odd.

This equation, being independent of ω_c, represents the envelope of the family of impulse responses obtained by varying ω_c from zero to infinity.

This observation leads to the following. Consider the impulse response when ω_c approaches infinity, that is to say for a network passing all frequencies without amplitude distortion and with a phase-shift of $-\omega t_0 \pm \frac{1}{2}\pi \pm 2m\pi$; or of $\frac{1}{2}\pi$ without taking into account $-\omega t_0$ and $2m\pi$, quantities which do not change the shape of the response of the network. The impulse response is always tangential to the envelope $E(t)$. As ω_c increases, the minima and maxima approach the vertical axis $t=t_0$. As ω_c tends to infinity, $2\pi/\omega_c$ tends to zero, and there is an infinite number of minima and maxima on the axis of abscissae and on the envelope $E(t)$, in any interval however small, so that it is practically impossible to discern the oscillations. It is, therefore, logical to replace this curve by one having the same element of surface between any two near ordinates, for this new curve will give the same value for the integral (7). That is to say, a 4-terminal network having this new impulse response will give the same response $R(t)$ as the preceding network, at least for frequencies not approaching infinity.

To find this new curve, we calculate the surface of the preceding one for $t_1<t<t_2$, t_1 and t_2 being so close to each other that t may be regarded as constant in the interval and equal to t_α.

We then have (supposing $t_0=0$)

$$\int_{t_1}^{t_2} I'(t)dt = \mp\frac{1}{\pi t_\alpha}\int_{t_1}^{t_2}(1-\cos\omega_c t)dt.$$

Integrating over one period, i.e. putting $t_2=t_1+2\pi/\omega_c$, we have

$$\int_{t_1}^{t_1+2\pi/\omega_c} I'(t)dt = \frac{2}{\omega_c t_\alpha}.$$

If $I_i'(t_\alpha)$ is the ordinate of the new curve, for $t=t_\alpha$, provided that the segments of the surface are equal, we have

$$I_i'(t_\alpha)\frac{2\pi}{\omega_c} = \mp\frac{2}{\omega_c t_\alpha}.$$

Hence, replacing t_α by t,

$$I_i'(t) = \mp\frac{1}{\pi}\cdot\frac{1}{t-t_0}. \qquad (18)$$

This equation represents the impulse of the network, i.e. the new form of the impulse when a 90° phase-shift is added at all frequencies. This curve is shown in (*a*), Fig. 4.

(*b*) *Band-pass filters.* From (15) we obtain directly, by integration,

$$I_B'(t) = \mp\left[\frac{\omega_{c2}}{\pi}\frac{\cos\omega_{c2}(t-t_0)}{\omega_{c2}(t-t_0)} - \frac{\omega_{c1}}{\pi}\frac{\cos\omega_{c1}(t-t_0)}{\omega_{c1}(t-t_0)}\right], \qquad (19)$$

an equation which represents the difference between the impulse responses of two ideal high-pass filters [see equation (21) below] whose cut-off frequencies coincide respectively with the higher and lower cut-off frequencies of the band-pass filter.

Formula (19) can also be written in the form

$$I_B'(t) = \mp\frac{\omega_{c2}-\omega_{c1}}{\pi}\frac{\sin\frac{1}{2}(\omega_{c2}-\omega_{c1})t}{\frac{1}{2}(\omega_{c2}-\omega_{c1})t}\sin\tfrac{1}{2}(\omega_{c2}+\omega_{c1})t. \qquad (20)$$

The shape of this curve is shown in (*b*), Fig. 4.

(*c*) *High-pass filters*. If the impulse response of an ideal high-pass filter is added to that of an ideal low-pass filter having the same cut-off frequency, the impulse response obtained, $I_i(t)$, is given by equation (18). Hence

$$I_H'(t) = I_i(t) - I_i'(t). \qquad (21)$$

The impulse response of the high-pass filter is then

$$I_H'(t) = \mp\frac{\omega_c}{\pi}\cdot\frac{\cos\omega_c(t-t_0)}{\omega_c(t-t_0)}. \qquad (22)$$

The shape of this curve is shown in (*c*), Fig. 4.

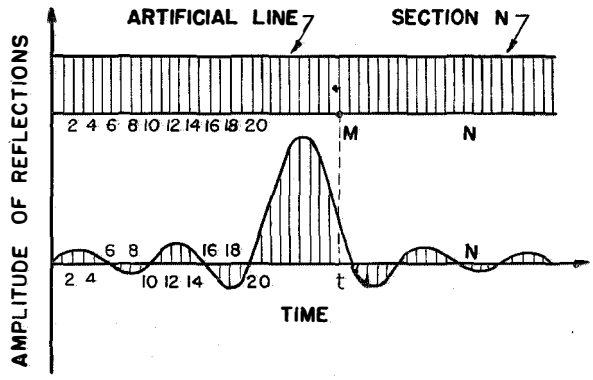

Fig. 6—Showing how the amplitudes of the reflections must be adjusted in order to obtain the required characteristics.

(3) *Applications*

(*3.1*) *Fundamental Principles*

One of the most important conclusions of the preceding study is that if two networks have the same impulse response, they are equivalent from the point of view of transmission.

Now, a very simple method exists of obtaining a given impulse response. This consists of using the circuit shown in Fig. 5, made up of a bridge driven at diagonally opposite points, the output voltage being taken off the other diagonally opposite points. There are 3 resistance arms, the fourth consisting of a real or artificial line. This line has a characteristic impedance Z_0 which is practically a resistance. The bridge is balanced with the line terminated in its characteristic impedance so that, if we apply a voltage to the input, the output voltage is zero. In particular, if an impulse is applied to the input, it will be transmitted to the line, propagated and absorbed by the terminal impedance.

To produce a definite impulse response, we proceed as follows. At equidistant points throughout the line we produce reflections, all of which are propagated towards the input of the line. A fraction will reach the output diagonal. The reflections will arrive at the terminals of the diagonal in succession at time-intervals equal to the time taken by the impulse to travel from one point where a reflection occurs to the next. If these intervals are very short, a continuous curve will be obtained and, by controlling the amplitude of each reflection, the desired shape of impulse response may be obtained.

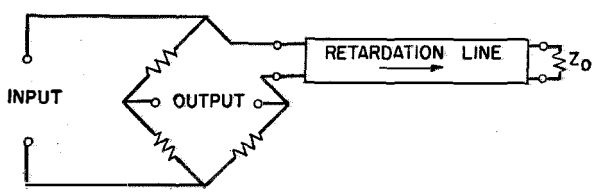

Fig. 5—Principle of a network incorporating an artificial line. By producing convenient reflections in the line any kind of characteristics can be obtained.

Let us, for example, construct a network having the impulse response shown in Fig. 6. Above this response is represented the artificial line of the circuit of Fig. 5. The scale of the diagram is such that if an impulse is applied to the input at time $t=0$, and if it arrives at the point M of the line at time t, the abscissa of this point on the diagram is also equal to t.

The line is divided into sections by a series of ordinates representing the points where reflections are produced. Two successive reflections are produced with a time-interval equal to Δt. Under these conditions, if we adjust each reflection to correspond in amplitude and size to the appropriate ordinate of the impulse response shown beneath the line, we shall obtain across the output diagonal of the bridge an impulse response of the desired form.

Naturally, when a reflection is produced, the line impedance and characteristics change, so that the propagation of the impulse along the line is no longer uniform; furthermore, multiple reflections may be produced. These difficulties are, however, avoided by producing reflections of very small amplitude, of the order of a few

per cent of the direct impulse. Of course, this results in a very small signal at the output and is a disadvantage of the system. However, this may be compensated by connecting an amplifier to the output.

From the above we conclude that with this circuit it is possible to construct a network having any desired characteristics.

A circuit of this type, which the author has constructed, and which, on account of the simplicity of adjustment, makes it possible to construct a compensating network or a circuit having a given characteristic in a few minutes, will now be described.

It will be observed that this method is particularly simple when it is desired to construct:

(*a*) A corrector circuit,
(*b*) A network equivalent to another one, or
(*c*) An electrical filter whose impulse response is known.

In the first case, it is sufficient to connect this circuit in series with the circuit which it is desired to correct and suitably adjust the reflections to correct the distorted impulse response; in the second, it is sufficient to obtain an impulse response identical to that furnished by the network to be copied. Finally, the given impulse response can be obtained quite simply.

The experimental arrangement will now be described and examples given of its application.

(*3.2*) *Experimental Arrangement*

This comprises (Fig. 7) an impulse generator, a reflecting network, and a cathode-ray tube circuit for observation of the impulse responses.

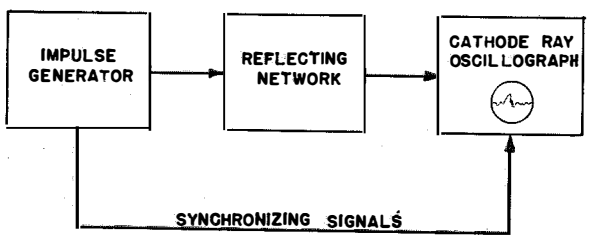

Fig. 7—Experimental arrangement for adjusting the network of Fig. 5. Any kind of characteristics can be obtained in a few minutes. An example is given in Figs. 12 and 13.

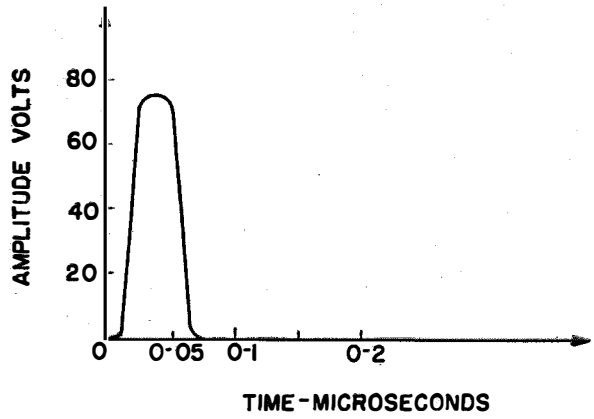

Fig. 8—Shape of the impulses produced by the generator.

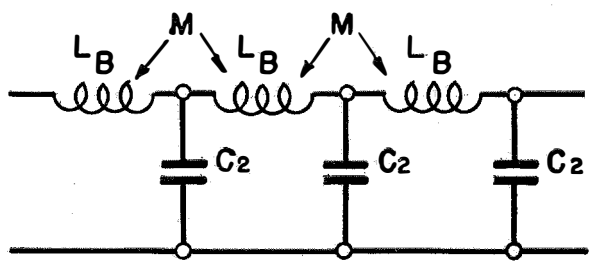

Fig. 9—Schematic showing some sections of the artificial line. The reflections are produced by varying slightly the capacitances C_2.

(*3.2.1*) *Impulse Generator.*

The impulse generator may be of any type. One may even (as we have done in the course of our study) use a square-wave generator. In this case the impulse responses must be replaced by indicial responses.

The impulses produced by the generator must be of very short duration, very much shorter than the period of the highest frequency to be transmitted. In the present state of technique, impulses having durations of less than 1 microsec. can be produced without great difficulty, and this makes possible the study of circuits at frequencies up to several megacycles per sec.

The author has developed a particularly simple circuit to obtain impulses of less than 0.1 microsec. duration at any desired frequency and having an amplitude of nearly 70 volts in a resistance of 300 ohms. The type of impulse is represented in Fig. 8. The harmonics produced by these impulses have a constant amplitude up to frequencies of the order of 15 to 20 Mc./s.

The peak power produced by these impulses is almost 15 watts, but the mean power is extremely small. A square-signal generator giving square signals of 15 watts peak will produce a mean power of some watts and will be much more complicated to design.

(3.2.2) Reflecting Network.

The reflecting network is of the type shown in Fig. 9, with input and output transformers.

The delay line is made up of a series of low-pass sections with mutual inductance between the successive sections (see Fig. 9).

The mutual inductance M is of such a value that the time-delay produced by the line is as nearly as possible independent of frequency. Theory shows that the best results are obtained in the following conditions.

If Z is the characteristic impedance of the line, and f_c the cut-off frequency of each section, the time-delay of the line will be constant within about 1% for all frequencies lower than $\frac{1}{2}f_c$, and the line elements are determined by the following formulae:

$$L_B=\frac{Z}{2\pi f_c}, \tag{23}$$

$$M=0.1L_B, \tag{24}$$

$$C_2=\frac{0.4}{f_c Z}, \tag{25}$$

with the notations of Fig. 9. The time-delay per section is given by the formula

$$T_1=\frac{0.4}{f_c}. \tag{26}$$

The reflections are produced by slightly detuning the capacitances of the line sections. By detuning two successive capacitances, two reflections are produced with a delay between them equal to T_1.

The number of sections necessary to reproduce a filter response suitably is very high. An estimate may be obtained by considering the impulse responses of Figs. 3 and 4. Except in the case of the low-pass filter of Fig. 4, it can be shown that the oscillations remote from the principal oscillation have little effect on the response curve of these filters, except at frequencies far removed from the cut-off of the frequencies. In practice, to reproduce, for example, the low-pass filter of Fig. 3, it is sufficient if only 3 secondary maxima to the left and 3 to the right of the principal maximum are retained. The width of the impulse response then corresponds to a duration of the order of $7(2\pi/\omega_c)=7T_c$, where T_c is the period of the cut-off frequency of the filter. If, on the other hand, we approximate each part of the curve comprised in an interval whose width is equal to T_c by 10 equidistant ordinates, the whole curve will be made up of about 70 ordinates. The retardation line should then consist of about 70 sections, the delay of each one being equal to $T_c/10$.

This number is large, but it can be achieved without much difficulty if the procedure indicated by M. Lalande[13] is followed. It consists of a continuous winding on a tube of insulating material, with equidistant shunt arms so placed as to obtain the self-inductance L_B of successive sections of the line. The relation between the self-inductance L_B and the mutual inductance M with the following element is obtained by suit-

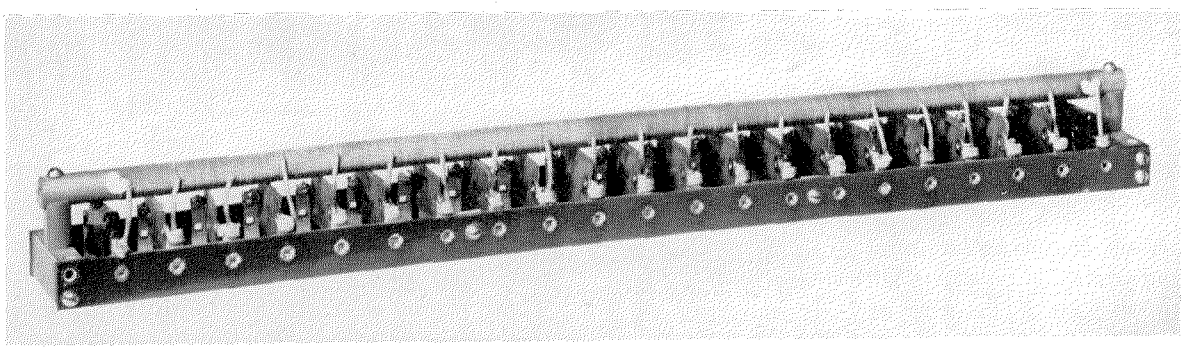

Fig. 10—Artificial line used in some practical filters.

ably choosing the diameter of the tube and the distance between the shunt arms. Lalande has shown that the desired relation is independent of the diameter of the wire used for the winding and is obtained if the distance between two shunt arms is equal to about 1.75 times the diameter of the tube.

A line constructed on these principles is shown in Fig. 10. It comprises 20 sections with a total time-delay of about 2 microsec. ($L_B \simeq 25\ \mu H$, $C_2 \simeq 170\ \mu\mu F$, $Z \simeq 600\ \Omega$ and $T_1 = 0.2\ \mu sec.$) The kind of filter working up to frequencies of the order of 2 Mc./s. can be made.

To adjust the amplitude of each reflection, each condenser C_2 must consist of a fixed value condenser in parallel with an adjustable trimmer. This is easily obtained by using trimmers of small dimensions with mica insulation and adjustable electrodes.

A model of this kind has been constructed by the author. It has 80 adjustable condensers and the whole circuit is contained in a box measuring $16 \times 10 \times 3$ in.

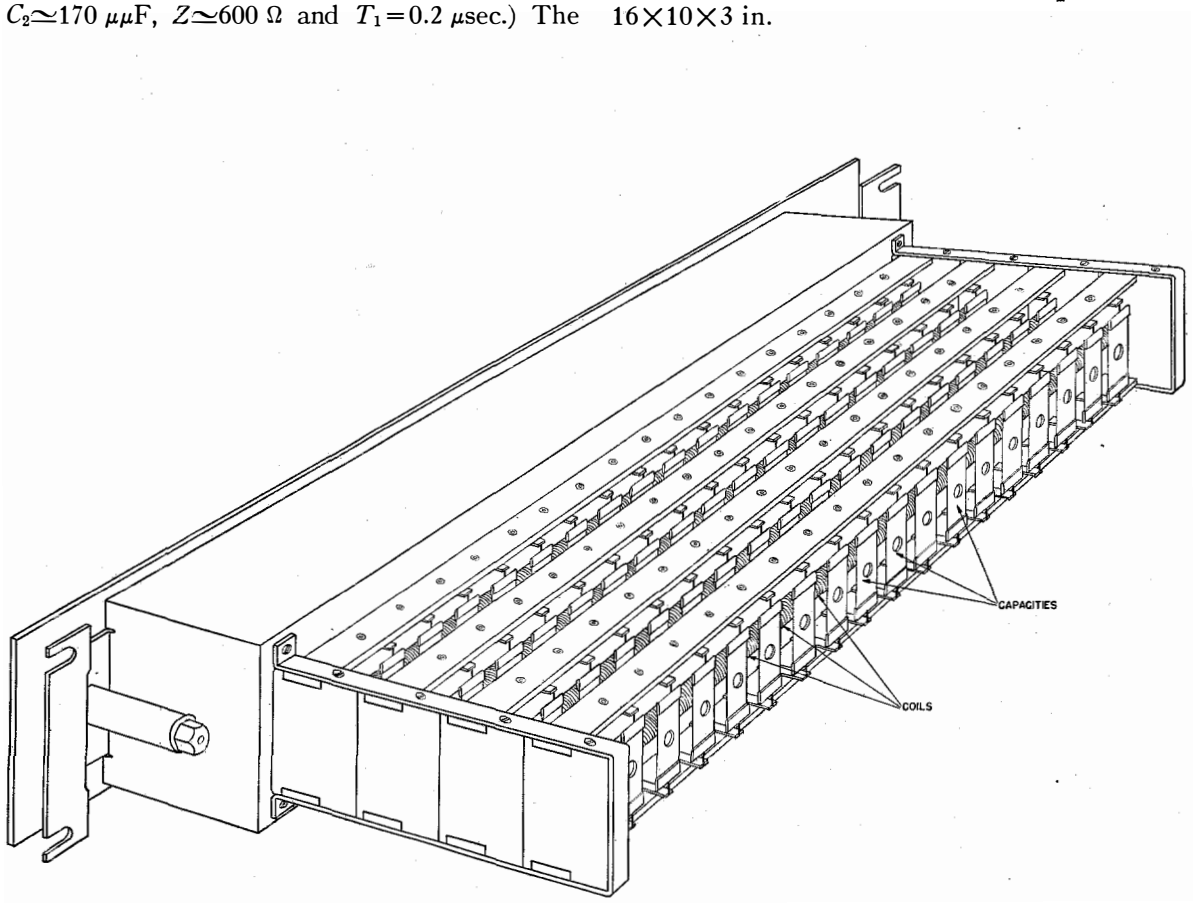

Fig. 11—Practical design of a linear phase-shift high-pass filter for television circuits.

cut-off frequency is equal to 4 Mc./s. and the time-delay is constant within nearly 1% up to frequencies of the order of 2 Mc./s.

Using 4 lines of this type, impulse responses defined by 80 points can be obtained, and any

(3.2.3) Experimental Results.

These results were obtained with the arrangement shown in Fig. 7. To facilitate the operation, the impulse response was traced on trans-

parent paper and fixed on the cathode-ray tube. By means of successive trimmers, the amplitude of each of the 80 successive ordinates was adjusted. This was easy since the ordinates follow in the same order as the trimmers throughout the length of the line. It has been found that any kind of impulse response can be obtained in a few minutes.

With this experimental arrangement, the author has made a series of filters and traced their amplitude and phase curves to see how they conform with those expected.

From this study the following experimental conclusions were drawn:

(*a*) The departures in shape from the theoretical impulse response do not greatly change the amplitude and phase curves. In particular the cut-off frequency depends solely on the pseudo-period of the oscillations of the impulse response and is practically independent of the amplitudes of these oscillations, at least for small departures (30% maximum) from the theoretical amplitudes.

(*b*) Between the pass-band and the attenuated band, attenuations of the order of 20–25 db. can easily be obtained. With a little more care (5 minutes' adjustment) 30 db. is reached without difficulty; but it is difficult to exceed this value by much.

(*c*) If it is desired to construct linear phase-shift filters and if the symmetry of the impulse response is adjusted by the eye alone, the phase curve is almost perfectly straight. It is easy to obtain good phase curves.

(*d*) On the other hand, good attenuation curves are more difficult to obtain. However, skill in adjustment is soon obtained, and then the desired modifications of attenuation can be effected.

The author has demonstrated with this experimental arrangement, and has passed from one kind of filter to another, completely different, in a few minutes.

He then decided to construct compact commercial filters with preset adjustments. One of these will now be described.

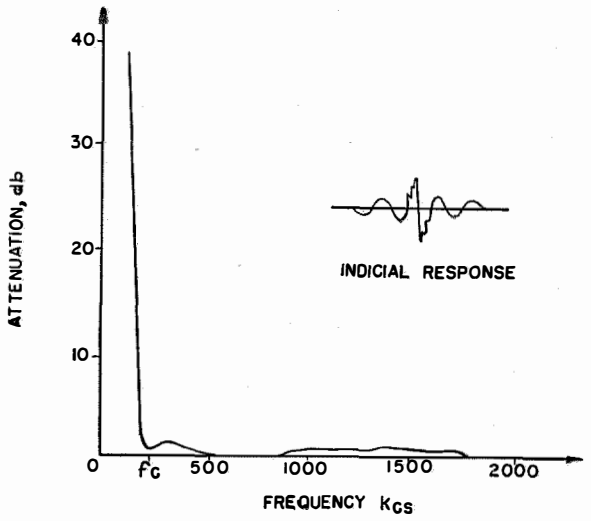

Fig. 12—Indicial response and selectivity curve of a practical linear phase-shift filter.

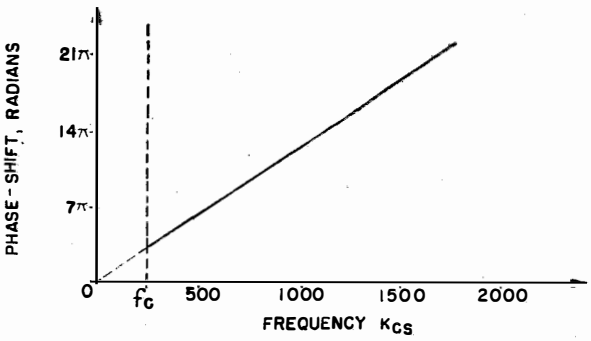

Fig. 13—Phase characteristic of the filter referred to in Fig. 12.

(*3.3*) *Typical Application*: *Filter for Television Circuits*

(*3.3.1*) *Filter Design incorporating Artificial Line.*

This filter is shown in Fig. 11.* It is a high-pass filter having a cut-off frequency of 250 kc./s. with a strictly linear phase-shift curve throughout the pass-band, the cut-off region, and even below into the attenuation band.

The artificial line is made up of 80 elements (4 lines, each of 20 elements) giving a delay of

* Reproduced from memory, as the original photograph is in enemy-occupied country.

7 microsec. from one end to the other, that is to say 14 microsec. for the total duration of the impulse response (delay between first and last impulses received at the output of the filter).

The capacitances are fixed and have values predetermined with the help of an identical filter with adjustable capacitances. It has been found that in order to obtain the balance of the bridge (no reflections in the line) the values of the capacitances must gradually increase from the beginning to the end of the line. This is due to the fact that the attenuation in each section gradually changes the characteristic impedance of the line. Slight readjustments may be made by scratching the surface of one of the silver layers through the windows to be seen in Fig. 11. These readjustments take only a few minutes and are very easily made with the calibrating circuit described above (Fig. 7).

The author is glad to be able to reproduce in Figs. 12 and 13 the true attenuation and phase characteristics of one of these filters.

When this filter was adjusted only a square-wave-generator was available and a suitable indicial response was produced on the cathode-ray tube. This curve (Fig. 12, inset) has been reproduced from a photograph. It will be seen that it is limited to three auxiliary oscillations on each side of the central peaks. This has been found sufficient to obtain a good attenuation curve up to more than 2 Mc./s. It may also be seen that the central oscillations contain some secondary peaks; these have been found to improve the attenuation characteristic at high frequencies.

A detailed study has been made by Boyer of the modifications which must be introduced into the theoretical impulse responses in order to obtain the best attenuation characteristic with the limited number of sections of the artificial line used in each case. It has been found that certain modifications of the impulse response (e.g. increase of the amplitude of some oscillations) produce a definite modification in the attenuation characteristic (sharper cut-off, for instance). By applying these observations, the indicial response of the above filter was produced and the attenuation characteristic of Fig. 12 was obtained.

The phase characteristic is linear (Fig. 13) and this was obtained without difficulty.

(*3.3.2*) *Comparison with the Linear Phase-Shift Filters of Bode and Dietzold.*

Compared with the filters of Bode and Dietzold,[8] the above filters present some advantages and some drawbacks.

Among the former, the artificial-line filters appear to be much easier to construct, are much more compact and require only elementary calculation. The phase curves are strictly linear and the loss in the attenuation band is comparable with that of the filters of Bode and Dietzold.

On the other hand, the artificial-line filters have the disadvantage—at least for the reflection models which have just been described—of having a loss of the order of 25–30 db. in the pass-band, while those of Bode and Dietzold have no appreciable attenuation in this band.

These conclusions were reached after the construction of two identical low-pass filters from the point of view of characteristics, one being constructed by the method of Bode and Dietzold and the other by the method described above.

The filter based on the method of Bode and Dietzold was calculated and constructed by a colleague of the author, Mr. Redard, and the artificial-line filter by the author (Fig. 11). During the comparative study the impulse responses of the two filters were compared and it was seen that the impulse response of the filter after Bode and Dietzold contained two or three secondary oscillations before the principal oscillation, and a great number of oscillations following the principal oscillation. This asymmetry explains the phase distortions which, though small, were still present in this filter.

The attenuation curves were nearly the same for the two filters.

It should be noted that it is possible to reduce the attenuation in the pass-band either by using stronger reflections—the adjustments are then more complicated, due to the multiple reflection—or by replacing one or more of the resistance arms of the bridge by an artificial line, each line introducing a gain of 6 db. in the pass-band.

(3.4) *Networks Producing 90° Phase-Shift at All Frequencies*

This kind of network differs from the preceding only in the form of its impulse response. It follows then that the circuits will be identical; but the characteristics obtained will be different because they are determined by the shape of the impulse response.

Before determining the practical possibilities, some essential properties will be referred to:

(*a*) The network will produce phase-shifts of 90° at all frequencies if the impulse response is odd with respect to its principal axis; as these odd-impulse responses are easy to obtain with the experimental circuit of Fig. 7, it follows that with the artificial-line circuit it is very easy to obtain networks producing a phase-shift of 90° at all frequencies ($\pm 2m\pi \pm \omega_0 t$).

(*b*) The difficulty appears when it is desired to produce an attenuation curve of a given shape, especially curves with a very sharp cut-off, for then the type of impulse response is also fixed, and if one cannot approximate it practically with a curve of relatively short duration, it will be necessary to use a line with a long transmission time and a large number of sections. Some examples follow.

(3.4.1) *High-pass Filters.*

The impulse response of this type of filter is shown in (*c*), Fig. 4. It is a curve with principal oscillations in the central region and oscillations decreasing progressively in amplitude to the left and right of this region. These secondary oscillations extend to infinity.

To reproduce this impulse response faithfully, an artificial line having an infinite number of sections is necessary. But, as in the linear phase-shift high-pass filter of Figs. 12 and 13, one may neglect the secondary oscillations and retain only the central part. This will modify the attenuation curve, which will no longer be the ideal one of (*c*), Fig. 4 but, as in the filter mentioned above, the characteristic may still be acceptable in practice.

It will be noticed that the oscillations of the impulse response, which extend to infinity, are similar to those of the linear phase-shift high-pass filter of (*c*), Fig. 3. In particular, the oscillations of these two impulse responses have the same envelope

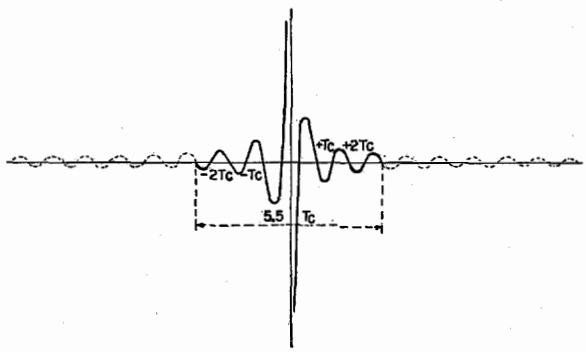

Fig. 14—Theoretical impulse response of a high-pass filter producing a phase-shift of 90° at all frequencies. The secondary oscillations shown in dotted lines can be neglected for usual applications.

$$E(t) = \pm \frac{1}{\pi(t - t_0)}$$

as may be deduced from equations (11) and (22).

It follows from this similarity that, as in the case of the above-mentioned filter, secondary oscillations may be neglected and, by confining the treatment to the central part, a sufficiently regular cut-off may be obtained.

As an example, let us limit the response to a duration $5.5T_c$, T_c being the period corresponding to the cut-off frequency of the filter. That is equivalent to neglecting the secondary oscillations shown in dotted lines in Fig. 14. If we approximate each wave of duration $\frac{1}{2}T_c$ by 10 equidistant ordinates, 110 are required in all. The artificial line will then need 110 sections—which are easily constructed—and a transmission time T from end to end equal to $5.5T_c/2$.

As an example, to make a high-pass filter of this type with a cut-off frequency equal to 100 kc./s. ($T_c = 10$ μsec.) one needs an artificial line having about 110 sections and giving a total delay of 27.5 μsec.

These filters will only produce the phase-shift of 90° and the predetermined attenuation characteristic for the frequencies at which the line is satisfactory. We have seen that the transmission delay of this type of line is constant to within approximately 1% for all frequencies less than $0.5f_{sc}$, where f_{sc} is the cut-off frequency of the sections of the line. Now, from (26) we have the

relationship

$$f_{sc}=\frac{0.4n}{T}$$

between the transmission delay T, the number of sections n, and the frequency f_{sc}. Hence, the line is satisfactory at frequencies less than

$$fc_2=0.5f_{sc}=0.2n/T$$

or, replacing T by $\frac{5.5Tc_1}{2}=\frac{5.5}{2fc_1}$

$$fc_2=\frac{0.4nfc_1}{5.5}\simeq\frac{nfc_1}{14}$$

$$\frac{fc_2}{fc_1}=\frac{n}{14}. \qquad (27)$$

As values of n between 100 and 200 are reasonable, it will be seen that it is possible to design high-pass filters of this type operating up to frequencies nearly 10 times higher than the cut-off frequency.

For instance, a high-pass filter of this type having a cut-off frequency of 100 kc./s. will give a phase-shift of 90° and good attenuation characteristics up to frequencies of about 1 Mc./s., if the retardation line comprises approximately 140 sections.

(3.4.2) Band-pass Filters.

The impulse response of this type of filter is shown in (*b*), Fig. 4. It is evident that to obtain a good attenuation characteristic the principal oscillation of the envelope and some secondary oscillations must be preserved. The duration of the impulse response is then proportional to the duration of each oscillation or fraction of oscillation of the envelope, i.e. to T_0 or $2\pi/(\omega_{c_2}-\omega_{c_1})$.

Thus, the duration is inversely proportional to the width of the pass-band. Consequently, for small band widths, lines of long delay and large physical dimensions are required. The larger the band width, the simpler is the design of band-pass filters of this type.

(3.4.3) Low-pass Filters.

The impulse response of this type of filter is shown in (*a*), Fig. 4. Here all secondary oscillations to the left of the vertical axis have negative amplitude, and those to the right positive amplitude. Consequently the value of the integral (17) is no longer negligible for these oscillations, especially at very low frequencies; the curve of attenuation will therefore be greatly distorted at these frequencies.

In practice it is impossible to design low-pass filters of this type with good transmission of frequencies in the neighbourhood of zero. The higher the lowest frequency to be transmitted, the simpler is the design of these filters.

The author hopes to present a more quantitative analysis of the preceding study in a subsequent paper.

(4) *Acknowledgments*

This work was carried out in Paris, at the Laboratories of Le Matériel Téléphonique in 1937–1939, and the author is glad to take this opportunity of expressing his sincere appreciation to the Technical Manager, Mr. E. M. Deloraine, and to Mr. Saphores, Chief Engineer, for their interest in the work. He wishes also to express his sincere appreciation to Mr. Van Mierlo, under whose supervision he worked for many years and to whom he is indebted for many useful suggestions.

The experimental work has been done by the author's friend and collaborator Mr. Boyer, and has given him a new occasion to admire Mr. Boyer's ability to overcome obstacles and his great skill in experiments.

The author has only one regret, that of being unable to give the beautiful intuitive and experimental conclusions of Boyer on the relationships between the modifications of the impulse response forms and the attenuation and phase curves. The greater part of the documents have unfortunately disappeared since the German invasion of France.

(5) *References*

Papers

(1) M. LEVY: "Transformations Sélectives. Application à l'Analyse des Mélanges de Sinusoides," *Comptes Rendus*, 1934, **198**, p. 2222.
(2) M. LEVY: "Nouvelle Méthode d'Analyse Spectrale des Courbes Non-périodiques," *ibid.*, 1934, **199**, p. 1031.
(3) M. LEVY: "Transformations Sélectives. Propriétés des Courbes de Transformations et des Courbes de Sélectivité," *ibid.*, 1934, **200**, p. 646.
(4) M. LEVY: "Théorie des Transformations Sélectives." Thesis and memorandum (Le Matériel Téléphonique), July, 1937.
(5) M. LEVY: "Application de la Théorie des Transformations Sélectives à l'Étude des Filtres Électriques," Memorandum (Le Matériel Téléphonique), 1937–1938.
(6) M. LEVY: "Étude des Distorsions Produits dans les Lecteurs de Films," *Revue Générale de l'Électricité*, 1935, **37**, p. 692.
(7) H. E. KALLMANN: "Transversal Filters," *Proceedings of the Institute of Radio Engineers*, 1940, **28**, p. 302.
(8) H. W. BODE and R. L. DIETZOLD: "Ideal Wave Filters," *Bell System Technical Journal*, 1935, **14**, p. 215.
(9) H. A. WHEELER: "The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes," *Proceedings of the Institute of Radio Engineers*, 1939, **27**. p. 359.
(10) J. R. CARSON and O. J. ZOBEL: Fundamental papers on the theory of wave filters, *Bell System Technical Journal*, 1922, 1923 and 1924.

Patents

(11) A. D. BLUMLEIN, H. E. KALLMAN and W. S. PERCEVAL: "Improvements relating to Wave Transmission Networks." British Patent No. 517516. Application date 28th June, 1938.
(12) N. WIENER and YUK-WING LEE: U. S. Patents No. 2,024,900 and 2,128,257.
(13) P. GLOESS and M. LALANDE: "Large-band Transmission System for Television." French Patent No. 845,402 (23rd April, 1938).
(14) P. GLOESS and M. LEVY: "Electrical Transmission and Correction Systems." French Patents Nos. 853,841 and 859,299.
(15) M. LEVY: "Electrical Transmission and Correction Systems." Patent of addition No. 509–26.

Marine Navigation Aids

The Radio Direction Finder and The Gyro-Compass

By E. H. PRICE, Manager, and W. J. GILLULE, Chief Inspector
Marine Division, Mackay Radio and Telegraph Company, New York, N. Y.

Editor's Note: The Radio Direction Finder, first introduced commercially by the Federal Telegraph Company, and the Sperry Gyro-Compass, have played and are playing a dominant role in marine navigation particularly aboard naval craft and merchant vessels. In this article the operation of the Radio Direction Finder and its automatic coordination with the gyro-compass are briefly indicated. The fundamental principles of the gyro-compass are then explained at some length inasmuch as its operation is seldom investigated or clearly understood by engineers in other fields. Further, it is felt that this exposition is of timely interest in view of great maritime expansion necessitated by war conditions.*

AS is well known, a Radio Direction Finder is a device making use of the directive receiving properties of a loop antenna to determine the direction of an incoming radio signal. On shipboard, the radio direction finder has become an important adjunct to navigation as it enables the navigator to determine the ship's position quickly and accurately in relation to radio beacons aboard lightships or at fixed shore stations. If the Radio Direction Finder were used as a simple homing device, the loop antenna could be made stationary with the plane of the loop athwart the ship's beam. This is, of course, not practical as it would limit the operation and usefulness of the instrument. Hence, the marine type Radio Direction Finder is equipped with a rotatable loop coupled to a pointer so that the *direction* of the received radio signal may be quickly indicated on an appropriate scale called the bearing or azimuth scale. (See Figs. 1, 2, and 3.)

If the Radio Direction Finder is installed on shore, the pointer indicates the source of a radio signal or direction on a bearing indicator scale permanently set to indicate the points of the compass. When the instrument is fitted aboard ship, the heading of the ship in relation to true north changes whenever the ship's course is changed so that a fixed scale would only indicate the direction of the received radio signal in relation to the beam of the ship and not its heading. The true direction of the radio signal might then be determined by interpolating the reading of the loop pointer on the fixed bearing scale against the true heading of the ship as indicated by the ship's compass.

In order to simplify this operation and prevent the possibility of error, the bearing scale on a ship's Direction Finder is made rotatable. Thus the bearing scale can be turned to indicate the ship's heading as read from a compass. If the bearing scale is properly set, the direction of the received radio signal may be read directly without complicated interpolation. In early Radio Direction Finders, the rotatable bearing scale could only be set by hand after the ship's heading had been verbally transmitted to the operator by the helmsman.

It was found advantageous, on ships equipped with Sperry Gyro-Compasses, to provide a repeater compass in close view of the Radio Direction Finder scale so that the person operating the Radio Direction Finder could accurately set the bearing scale from the repeater compass scale. Thus the cooperation of a second party was not required.

Both these methods had their disadvantages. When using a regular compass, only the sense of vision is required but, when using a Direction Finder, the operator depends both upon hearing and vision. If the bearing must be set by hand, the operator must first set the scale to the ship's course and then correlate the sound input of the receivers with the ship's heading and note the reading on the scale. Before this can be done, the ship may have yawed several degrees off its course, appreciably affecting the results.

Soon after the Federal Telegraph Company perfected the first commercial marine Radio Direction Finder in 1921, its engineers began de-

* Now the Federal Telephone & Radio Corporation.

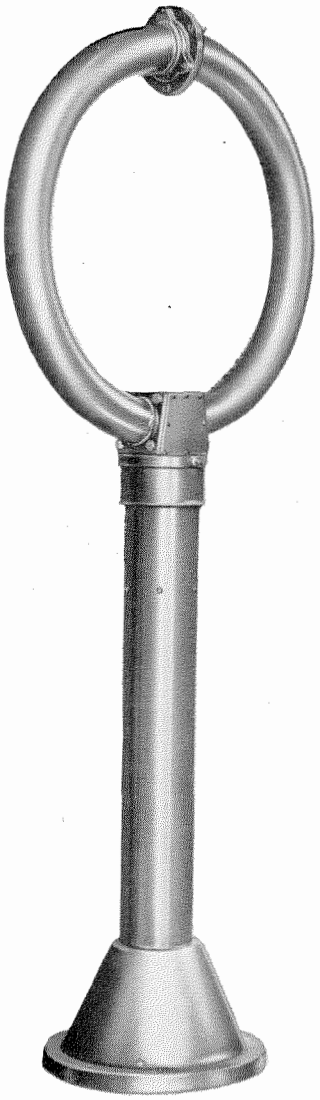

Fig. 1—A modern rotatable loop manufactured by the Federal Telephone & Radio Corporation for use with its marine radio direction finders. The loop, as shown, is installed above the deck of the ship.

velopment work on a Direction Finder equipped with a rotatable bearing scale automatically indicating the ship's heading. A Radio Direction Finder was designed that could be equipped with a Sperry Gyro-Compass repeater motor to turn the bearing scale. Further, engineers of the Sperry Gyroscope Company cooperated by designing a repeater motor for this purpose (Fig. 4). The resultant Radio Direction Finder operates with far greater accuracy as the radio bearings then can be read directly and instantaneously once the Direction Finder repeater motor is synchronized with the ship's gyro-compass. The first Radio Direction Finder equipped with a Sperry Gyro-Compass repeater motor was installed by Federal on the S.S. Manhattan in the fall of 1932. Today more than fifty percent of the Radio Direction Finders installed on ships use this repeater feature.

The importance of the gyro-compass and the Radio Direction Finder has been enhanced by their coordination. In view of the fact that the operation of the gyro-compass is not generally understood by engineers not in the marine field, the following description is felt to be of timely interest.

THE GYRO-COMPASS *

The word "gyroscope" is of French origin—a combination of two Greek words, "gyros," meaning turn or revolution, and "skopein," meaning to view, the literal translation of the two words being "to view the revolution" of the earth. The correct pronunciation of the word is with the g soft as in "gentleman." In the first syllable the y is long, as in "sky." Ro is pronounced like the row in rowboat, and scope to rhyme with rope.

The reason the French have the distinction of originating the name gyroscope is because the great French scientist Leon Foucault was one of the first authorities on the subject of gyroscopic phenomena, having succeeded as early as 1852 in actually producing a gyroscope with which he could observe, with the aid of a microscope, the ceaseless onward movement of the earth's rotation.

If the reader will permit one more slight digression it might be well to remember (before we get down to business) that the sphere on which we live is in itself a mammoth gyroscope and that there would be no life at all on the earth if it did not revolve like a top, with the direction of its polar axis fairly constant. Otherwise the surface of the earth would be exposed to extremes of heat and cold with such rapidity that living organisms would not be able to survive.

* Extracted from a manual on the Sperry Gyro-Compass and Gyro-Pilot, published by Sperry Gyroscope Company, Inc., Brooklyn, New York, and reproduced by permission.

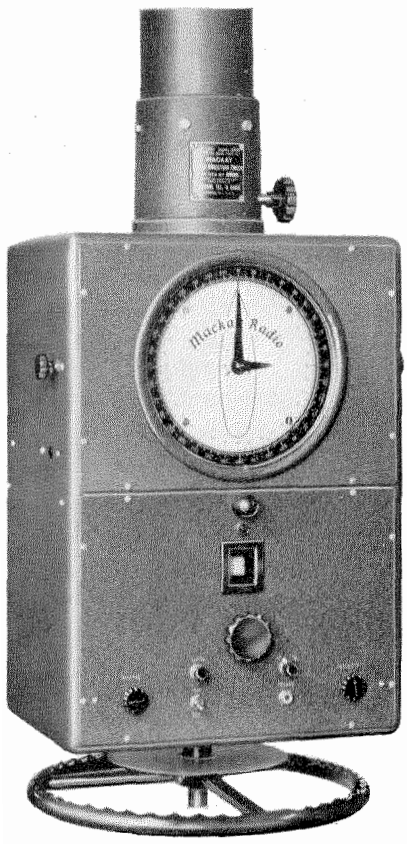

Fig. 2—Type 106 Radio Direction Finder manufactured by the Federal Telephone and Radio Corporation.

Definition and Principles of the Gyroscope

There is nothing mysterious about the gyroscope. Its actions, though they may appear at first to defy the laws of physics, in reality depend entirely upon Sir Isaac Newton's Laws of Motion.

Many of the toys we used to play with were based on gyroscopic principles. A spinning top is an elementary form of gyroscope, the "diabolo" —once a popular object for pastime—is another; so also is a hoop, for it will exhibit the characteristics of a gyro as long as it has sufficient motion to roll along the ground.

All of the practical applications of the gyroscope are based on two fundamental characteristics, namely: "Gyroscopic Inertia" and "Precession."

Gyroscopic Inertia, or rigidity in space as it is sometimes known, is the tendency of any rotating body to preserve its plane of rotation. For example, a hoop, when set in motion, will keep on rolling if undisturbed, approximately in a straight line, instead of tipping over as it would if not revolving.

The second characteristic of the gyroscope—Precession—is also exhibited by the rolling hoop. If we wish to change its direction of travel, we do not press against the rim at the front or back, but at the *top*—as though we intended to tip it over about an imaginary horizontal axis. The hoop resists this pressure and turns, instead,

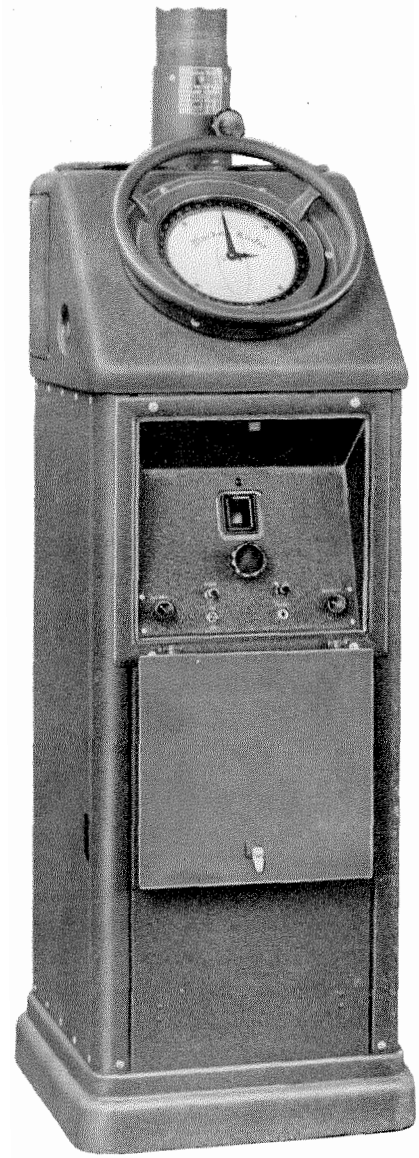

Fig. 3—Type 105 Radio Direction Finder manufactured by the Federal Telephone and Radio Corporation.

about a vertical axis which is at right angles to the axis about which the pressure was applied.

If we transform the hoop into a wheel, provide an axle for it, and mount the axle in supporting rings as shown in Fig. 5, we can obtain a true gyroscope, which is simply a spinning wheel or mass, universally mounted. Only one point—the geometrical center of its supporting system—is in a fixed position, the wheel being free to turn in any direction around this point. The wheel or rotor is free to revolve in its supporting ring about axis 1. The supporting ring is free to revolve in an outer ring about axis 2 which is always at right angles to the axis of rotation of the wheel. The outer ring, likewise, is free to revolve in pivot bearings in a supporting frame about axis 3 which is always at right angles to the axis of rotation of the inner ring.

With this arrangement, the axle can be pointed in any direction without altering the geometrical center of the assembly. When such a wheel is spinning, it exhibits exactly the same characteristics as the hoop, but does so without having to be rolled along the ground. "Gyroscopic inertia" may be illustrated by spinning the rotor and placing it in the position shown in Fig. 6. If the base of the gyroscope is tilted, as shown in Fig. 7, the rotor, instead of tipping over, as it would if not revolving, maintains its original plane of rotation. It will continue to do so, no matter how much the base of the gyro is moved about, as long as it continues to spin with sufficient velocity to overcome the friction between itself and its supporting bearings.

This characteristic is the result of the action of forces affecting the state of rest and motion of a gyroscope in the manner expressed by Newton's First Law of Motion, which states that *every body continues in its state of rest or of uniform motion in a straight line, unless it is compelled by forces to change that state.* This law as applied to a rotating wheel may be expressed by stating that a rotating wheel tends to maintain the direction of its plane of rotation in space and the direction of its axis in space.

"Precession" may be illustrated by applying a force or pressure to the gyro about the horizontal axis as shown in Fig. 8. It will be found that the applied pressure meets with resistance and that the gyro, instead of turning about its horizontal axis, turns, or "precesses" about its vertical axis in the direction indicated by the arrow P. Similarly, if we apply a pressure about the vertical axis, the gyro will precess about its horizontal axis as shown at P in Fig. 9. If there were a complete absence of inertia and friction about the precessional axis, the rate of precession would be such that the resistance of the gyro would be exactly equal to the applied pressure at any instant, and no movement from this pressure could ensue until the gyro had precessed so that its plane of rotation coincided with the plane of the applied pressure. Then the precession would cease and, with it, all resistance to the applied pressure.

A convenient way to remember the direction in which precession takes place is to regard the pressure as though it acted at a single point on the rim of the wheel, as indicated by the black dot in Fig. 8. This point will not move, in response to the pressure, but a point of 90 degrees beyond, in the direction of the wheel's rotation, will move away instead.

There you have the gyroscope in a nutshell, but inasmuch as the next step will be an explanation of the reason for precession, you may wish

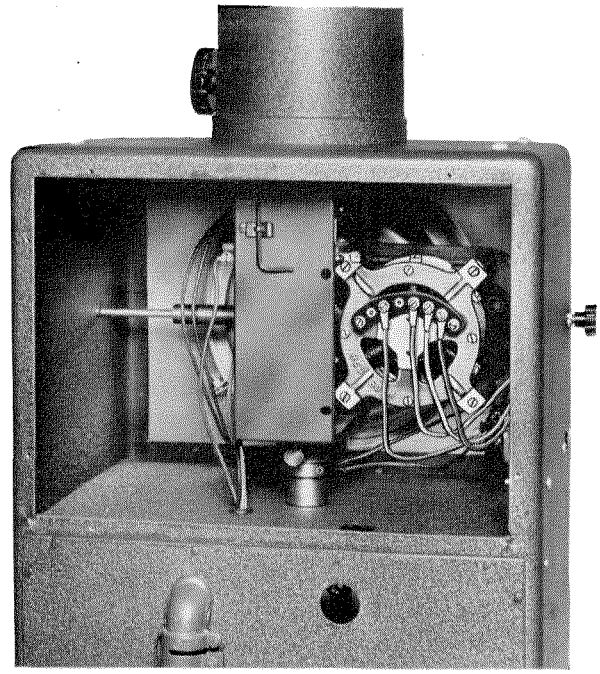

Fig. 4—Rear view of a Type 106-Federal Radio Direction Finder equipped with a Sperry repeater motor to rotate the bearing scale in relation to changes in the ship's heading.

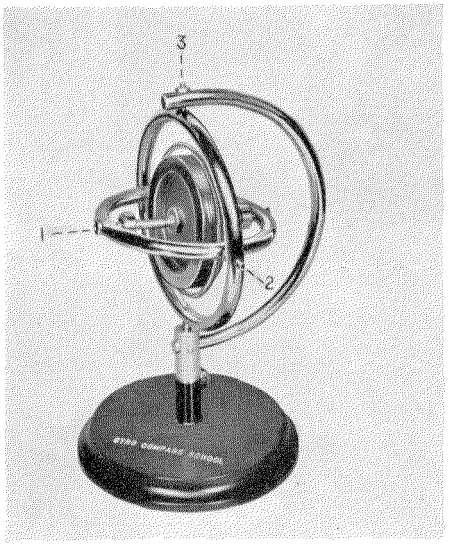

Fig. 5—The gyroscope has three axes of angular freedom.

to ask a question or two at this point. You might like to ask, for instance, why it is that of the many rotating objects with which you are familiar, some display gyroscopic phenomena and others do not. Why does a *top* exhibit gyroscopic characteristics while an engine flywheel, which also spins with high angular velocity, does not? Why does a *rifle bullet* behave like a gyroscope while a *windmill* merely behaves like a windmill?

Gyroscopic properties are inherent in all rotating masses, but can best be observed in those which have the greatest amount of freedom about two axes in addition to the spinning axis. The top comes under the latter category. The engine flywheel, on the other hand, is limited to one angular axis of freedom—its spinning axis. A rifle bullet may be likened to a gyroscope because it is free to revolve about two other axes, in addition to its spinning axis; therefore it exhibits gyroscopic inertia, tending to maintain a straighter line of flight than it would if not rotating. A *windmill* has freedom about its spinning axis and also about a vertical axis (as it must be able to turn in any direction under the control of its rudder). It has no freedom about a horizontal axis other than its spinning axis, however, and therefore, although precessional forces are impressed upon the apparatus by shifts of wind, there are no visible effects. The precessional forces result in a torque which is absorbed in the bearings. In a windmill these forces are small, however, owing to the light construction of the fan. In order to obtain maximum gyroscopic effects a rotor should be comparatively heavy, with as much of its weight concentrated at the rim as practicable, and it should spin with considerable velocity. Gyroscopic inertia depends upon angular velocity, weight and radius at which the weight is concentrated. Maximum effect is obtained therefore from a mass, with its principal weight concentrated near the rim, rotating at high speed.

Fig. 6—When spinning, the gyro exhibits "gyroscopic inertia."

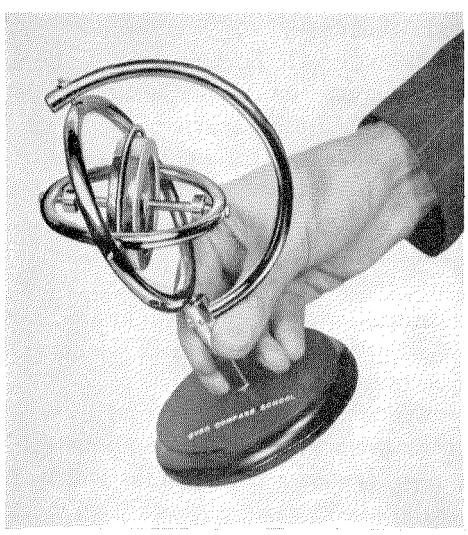

Fig. 7—The original plane of rotation is maintained no matter how the base is moved about.

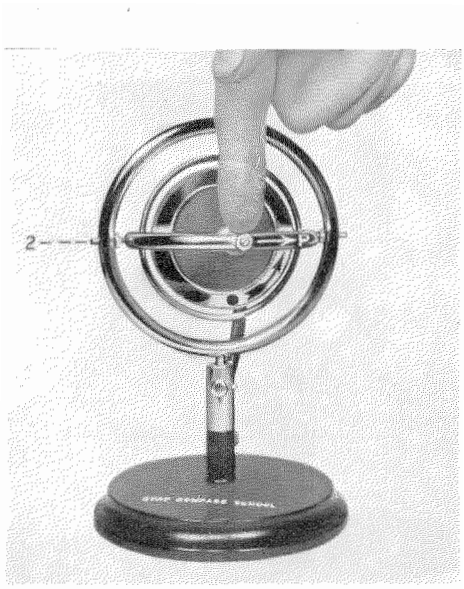

Fig. 8—Precession about the vertical axis.

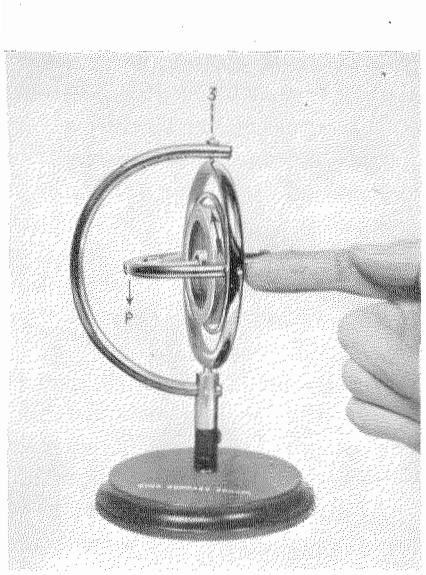

Fig. 9—Precession about the horizontal axis.

Reason for Precession

The reason for precession may be explained quite simply if we consider the mass of the rotor to be concentrated in separate particles such as A, B, C, and D in Fig. 10. The latter is a section through the center of the rotor, just as though you sliced it in half with a knife, threw the upper half away and lay the bottom half on the paper. We will assume that the wheel is spinning with considerable velocity in the direction of the arrow R at the top, and we will select that instant in the cycle when the particles are in the positions shown in the figure. If we can show what will happen to four particles equally spaced as are A, B, C, and D, we can show what happens to the entire rotor, since all other particles within it act in the same manner.

In order to make the explanation clearer we will simulate the movements of the rotor by corresponding motions of this book itself, and as a first step in this direction we will lay the book flat upon the table.

Now let us assume that a force F is applied against the rotor just as though we pressed down against the paper with a pencil at this point—as though we tipped the top of the book *down*, the bottom *up*. This would tend to rotate the wheel about the axis X–X′. Sir Isaac Newton said, in effect, that all matter is pigheaded or stubborn—that it will continue to move in a straight line unless disturbed, and if disturbed, it will offer resistance to the disturbing force. Let us see what happens to particle A. This particle was moving to the right before we started pushing down at F. Now however, it tends to move to the right and *down* into the paper—a combination of the motion due to the wheel's rotation and the motion due to our applied force F. Likewise particle C, which was moving to the left, now tends to move to the left and *up* out of the paper.

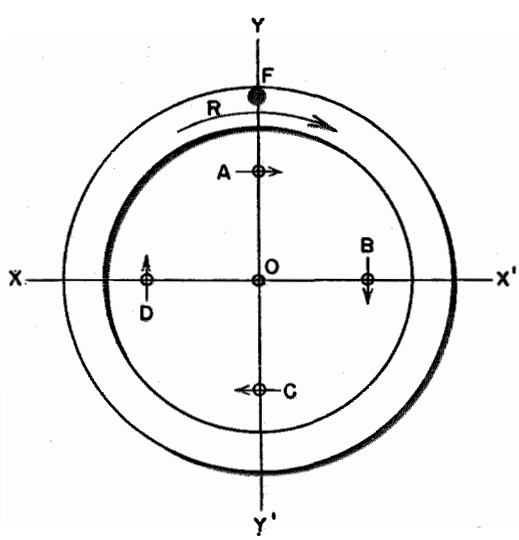

Fig. 10—Gyro rotor shown in section. For the sake of clarity its mass is considered to be concentrated in the four separate particles, A, B, C, and D.

As a result of these motions the wheel actually turns about the axis YY′ which is at right angles to the axis about which the force F was applied. Its motion is the same as though we tipped the right hand side of the book *down*, the left *up*. This is precession.

The reader will probably ask: "What happens to B and D?" Since B and D lie in the axis about which the force F is applied, they are unaffected by that force. Like A and C, however, they are pigheaded, and want to have their own way. Because of the wheel's rotation, B moves toward the bottom of the page, D toward the top. But the wheel is now turning about axis YY′ because of its precession. Therefore B tends to move toward the bottom of the page and *down* into the paper, D tends to move toward the top of the page and *up* out of the paper. In a perfectly balanced gyro operating without friction, the sum of these motions would exactly offset the force F, so that no motion could take place about axis XX′. Thus the only motion which could result from the application of a force as at F would be precessional rotation about an axis at right angles to the axis about which the force is applied. In other words, the wheel moves in the direction of the least resistance to any force which tends to disturb its plane of rotation—and the point of least resistance is always 90 degrees away in the direction of the wheel's rotation.

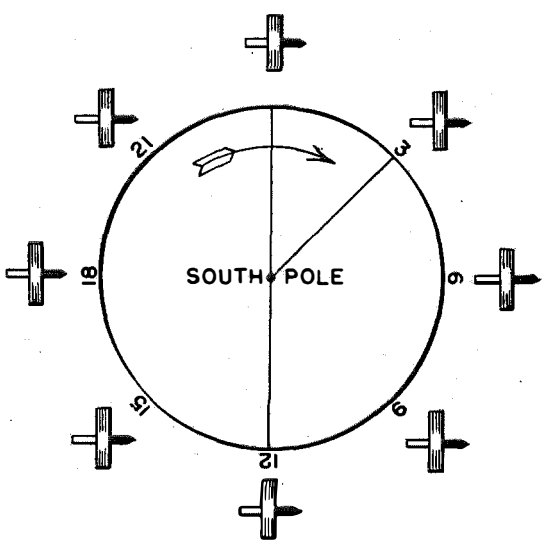

Fig. 11—A gyro with its spinning axis set in the East-West position at the equator appears to turn about its horizontal axis once each twenty-four hours.

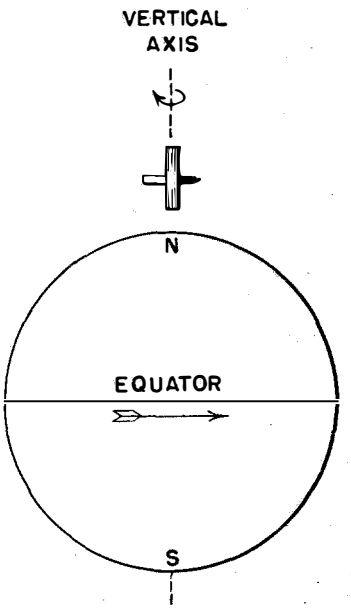

Fig. 12—A gyro with its spinning axis set horizontal at the Pole appears to turn about its vertical axis once each twenty-four hours.

Operating Principles of the Gyro-Compass

In the Gyro-Compass the characteristics of the gyroscope, "inertia" and "precession," which we have just explained, are combined with two constant, natural phenomena—the earth's rotation and the force of gravity, with the result that the instrument aligns itself with the *geographic* meridian and provides a constant true north indication regardless of the rolling, pitching, and yawing of the vessel.

Apparent Rotation

Let us consider the gyro to be mounted at the equator with its axle east and west. We will observe its behavior from a point in space beyond the South pole as shown in Fig. 11. To avoid confusion we will dispense with the supporting rings in this and subsequent illustrations, and show only the wheel and axle of the gyro, as these are the parts with which we are concerned chiefly.

The earth turns in the direction of the arrow, or clockwise, with an angular velocity of one revolution every 24 hours, carrying the gyro around with it; but the gyro, because of its inertia, maintains its original plane of rotation in regard to *space* just as it did when its base was

tilted as shown in Fig. 7. With respect to the *earth*, however, the gyro rotates about its horizontal axis with an equal velocity (one revolution in 24 hours) but in the opposite direction to the rotation of the earth. After three hours the end of the axle which was pointing east apparently is elevated at an angle of 45 degrees; after six hours, 90 degrees; after twelve hours, 180 degrees; and so on, until, at the end of 24 hours, it is back where it started.

Similarly, if we consider the gyro to be placed either at the North or the South pole at the theoretical axis of rotation of the earth, with the axis of the gyro horizontal, as shown in Fig. 12, the gyro will appear to rotate, but this time about its vertical rather than its horizontal axis.

At points between the poles and the equator the gyro appears to turn partly about the horizontal axis and partly about the vertical, because it is affected by both the horizontal component and the vertical component of the earth's rotation (see Fig. 13). The horizontal component of the earth's rotation causes the north end of the axle to rise. The vertical component causes it to turn to the east.

The reader will perceive that the difference between gyroscopic inertia and apparent rotation is simply one of *point of view*. As far as *space* is concerned the gyro remains fixed. In comparison with the earth, however, the gyro actually rotates as described above. It is this rotation which makes it possible to apply the force of gravity so as to convert the gyroscope into a North seeking gyro-compass.

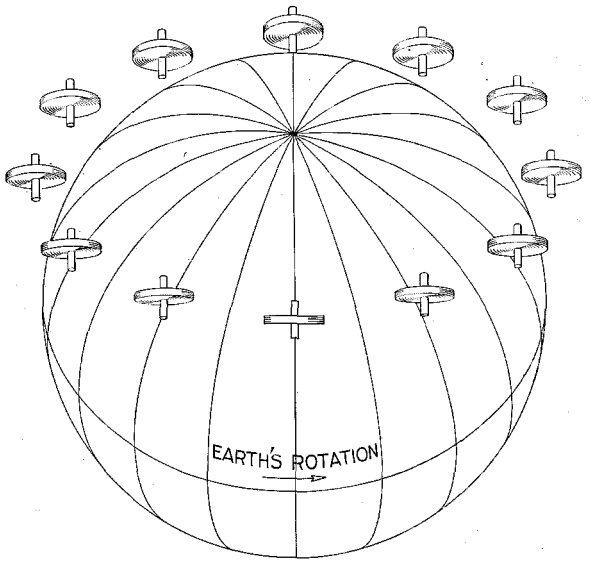

Fig. 13—A gyro with its spinning axle set horizontal at any point away from the equator maintains its plane of rotation in space and apparently moves about both its horizontal and vertical axis.

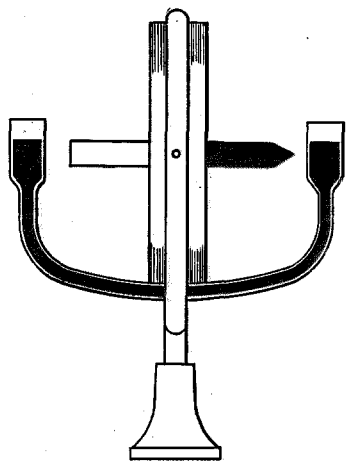

Fig. 14—To make the gyro seek *the North a mercury tube is added, its effect being applied about the horizontal axis.*

The first step in this direction is to cause the gyro to precess toward the meridian. Fig. 14 shows diagrammatically a gyro to which has been added a pair of containers with interconnecting tube; the assembly is partially filled with mercury and is attached to the gyro frame in such a way that it will tilt with the gyro when the gyro tilts or rotates about its horizontal axis. With the gyro at the equator and horizontal as shown at A in Fig. 15, the mercury is distributed equally in the tube and its weight exerts an equal downward pressure on each end of the axle. Therefore, in this position, the mercury has no effect upon the gyro. As the end of the axle which is pointing east (the right-hand end) slowly rises, some of the mercury, under the influence of gravity, is transferred to the lower end of the axle, as shown by the arrow at the left in B of Fig. 15. In this position a force is being exerted about the horizontal axis; the effect of the mercury being the same as though we were to push down on the west end (the left-hand end) of the gyro axle. The result is that the gyro precesses about the *vertical* axis as shown by the small arrows at the top in C and D, the axle turning slowly counter-clockwise. As the end of the gyro which at first was pointing

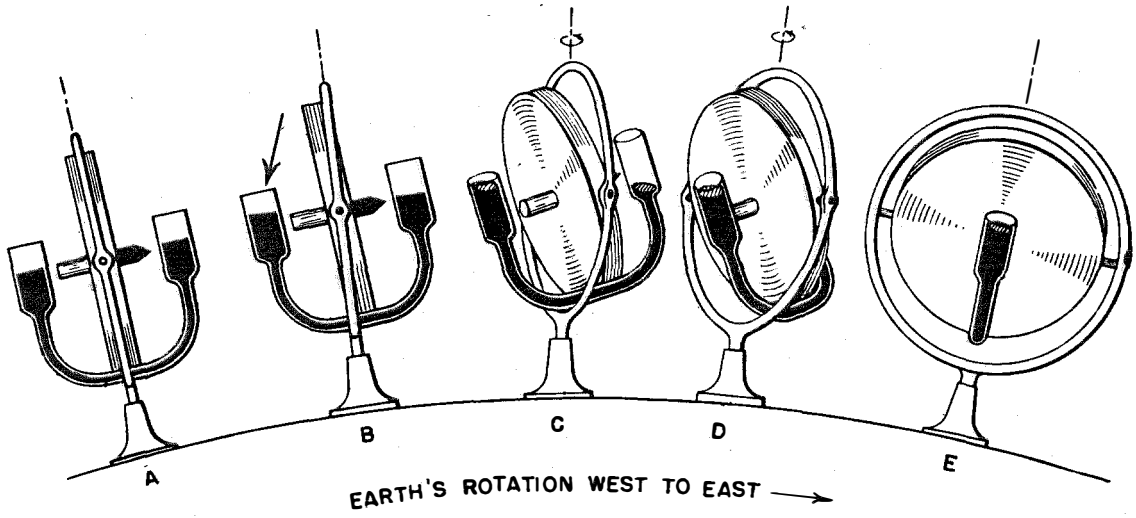

Fig. 15—Effect of the mercury ballistic when applied about the horizontal axis.

east (which we shall now refer to as the north end) precesses toward the meridian, it continues to rise, under the influence of the earth's rotation. After the gyro axle has precessed so that it is parallel to the meridian, the excess mercury at the south end causes its precession to continue, and the end of the gyro axle which was at first pointing west (which we shall now refer to as the south end) is carried to the east of the meridian. This south end now commences to rise and the mercury starts flowing back to the north end, precession being gradually diminished until the axle is again horizontal and the mercury evenly distributed. At this point precession of the north end toward the west ceases. The south end continues to rise, however, because it is still east of the meridian, and at length the mercury in the north side of the tube overbalances that in the south side. Precession, therefore, is reversed, and the north end returns toward the meridian, declining more and more as the south end continues to rise. This oscillation of the gyro about the meridian may be clearly understood by referring to Fig. 16 which shows the movements of the gyro axle projected onto a vertical plane. The ellipse in Fig. 16 is the result of a displacement of the gyro axle of only a few degrees from the meridian. If the gyro axle were pointing east and west at the beginning of the cycle, as shown in Fig. 15, precession would take place through 180 degrees in each direction, and at one extreme the north end of the gyro axle would point east, at the other, west. In any case the gyro never comes to rest, as there is no force tending to restore its axle to a horizontal position until after it has passed the meridian.

The ratio of the movement about the horizontal axis (caused by "apparent rotation") to the precessional movement about the vertical axis (caused by the flow of mercury) determines the shape of the ellipse. If the free surface of the mercury in the containers is increased so that more mercury can be transferred, the rate of precession will increase and therefore the ellipse will be flatter. If the mercury effect is decreased, the rate of precession will diminish to a point where the ellipse would, theoretically, be almost circular.

In the preceding paragraphs we have explained the behavior of a gyro under the simplest form of mercury control, the mercury being attached directly to the ring, frame or casing which supports the gyro. With such an arrangement the mercury can act only about the horizontal axis, and the gyro, therefore, will precess only about the vertical axis. An additional pressure is required about the vertical axis in order to generate precession about the horizontal axis that will counteract the natural tendency of the gyro axle to tilt. The manner in which this is accomplished will be shown in the following paragraphs.

It will be necessary first, however, to explain

the basic elements of an actual gyro-compass. As shown in Fig. 17 the rotor is contained in a case (1) and the case is supported on horizontal bearings in a vertical ring (2). The rotor-case and the vertical ring are free to turn about the vertical suspension axis (3). Although the gyro-compass, as shown in Fig. 17, necessarily differs in its details of construction from the model gyros shown in some of the previous illustrations, it has the same angular freedom about its spinning, horizontal, and vertical axes, and exhibits exactly the same characteristics.

Fig. 17 shows the addition of an outer frame (4) called the phantom, which is driven by an electrical follow-up system so that it follows every movement of the gyro about the vertical axis. By supporting the mercury tube or ballistic (5) in bearings in the phantom ring we can obtain a controlling action about the vertical axis of the gyro so as to arrest its oscillations and cause it to align itself with the meridian.

This is accomplished by connecting the mercury ballistic to the gyro case at a point (6) *slightly to the east of the vertical centerline.* With this arrangement the major effect of the mercury still acts about the horizontal axis and causes the gyro to precess toward the meridian as before; but there is now an additional effect about the vertical axis which causes the gyro to precess about the horizontal axis, introducing a tilt of the gyro counter to the natural tilt resulting from "apparent rotation." The end of the axle will therefore follow a spiral path as shown in the polar diagram, Fig. 18. The reduction of the oscillation produced by the action of the mercury

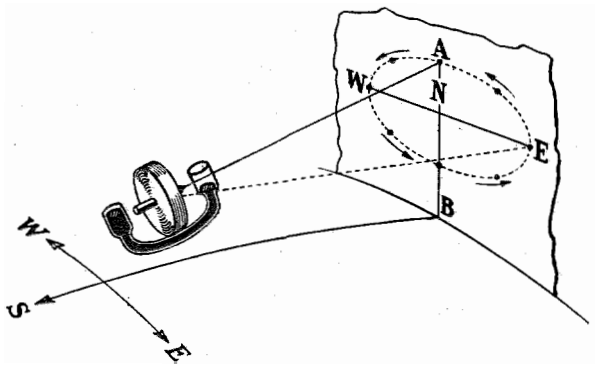

Fig. 16—Diagram showing the movement of a mercury-controlled gyro wheel when set with its axle pointing east of north.

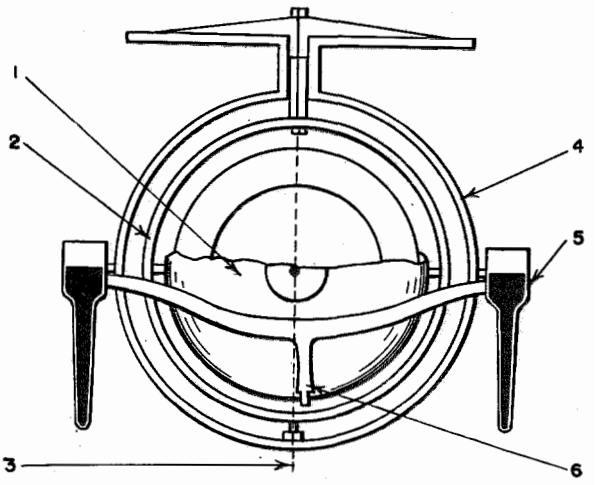

Fig. 17—Elements of the Gyro-Compass. In order to obtain a more symmetrical construction, the mercury ballistic consists of two sets of containers and tubes, instead of the single pair of containers shown in preceding illustrations.

ballistic about the vertical axis is called "Damping." A careful consideration of the action of the mercury ballistic will make it apparent that the only position of rest which the gyro can find *at the equator* will be with its axle horizontal and in the meridian. In other words, we have obtained a true, meridian-seeking Gyro-Compass.

A number of other factors must be considered, however, before we can obtain a gyro-compass which will function accurately and reliably, at various latitudes, on a rolling, pitching vessel moving over the earth's surface at considerable speed.

We have seen that the action of the mercury ballistic about both the horizontal and vertical axes is made possible by the use of the phantom element. This element serves another important purpose: it provides a means of suspending the gyro so that it is practically free from friction about its vertical axis. The gyro is supported from the top of the phantom by steel wires and the phantom is kept in exact alignment with the gyro by means of an electrical follow-up system. The compass card is a part of the phantom element, the whole of which is supported on ball bearings from the main supporting frame or "spider." Thus, with the exception of the eccentric connection between the mercury ballistic and the gyro case and the upper and lower guide bearings, which are practically frictionless, there is no

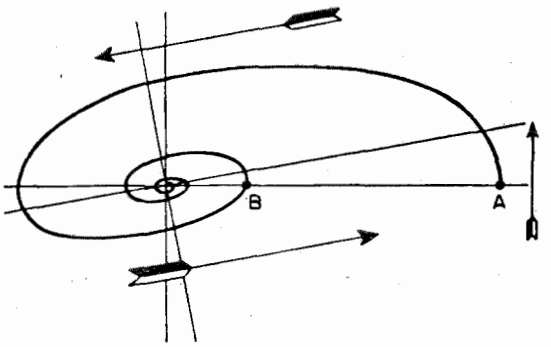

Fig. 18—Action of the gyro axis when the mercury ballistic is connected to its casing through an eccentric pivot.

physical contact which can result in any friction between the sensitive gyro element and the compass card or other external parts.

If the compass were to be used on shore, it would be feasible to control the gyro simply by suspending a weight from the phantom, like a pendulum, and connecting the weight to the eccentric pivot on the bottom of the rotor case. This would be impractical on board ship, however, where a compass is subjected to rolling movements in an intercardinal plane (northeast to southwest, or northwest to southeast). Intercardinal rolling causes a compass to swing in its gimbals, with the result that the pendulum would be subjected to acceleration forces which would cause a continuous torque about the vertical axis of the compass.

One way of avoiding this effect would be to stabilize the compass gyroscopically and so prevent it from swinging. In the Sperry compass, however, the complication of stabilizing gyros is avoided by the use of the mercury ballistic, which controls the gyro as we have already explained. The mercury ballistic is non-pendulous; its weight is distributed equally above and below the gyro axle so that it is neither top-heavy nor bottom-heavy. Therefore, no acceleration forces are generated and no torque about the vertical axis of the compass is introduced by the swinging of the compass in its gimbals.

Under slow rates of inclination such as those produced by the earth's rotation, the action of the mercury ballistic is equivalent to that of an ordinary pendulum, but opposite in direction. Under rapid rates of movement, however, the small bore of the mercury tubes prevents the mercury from surging back and forth and introducing errors in the compass.

Damping Factor

The extent of the damping action is governed by the displacement of the mercury ballistic connecting link from the centerline. Commercial compasses are given a damping factor of 66%, i.e., the eccentricity of the connecting link is such that each swing of the gyro axle from the meridian is one-third of the preceding swing, the amplitude being reduced by 66% at each oscillation. If the compass is started 30 degrees east of the meridian, the first swing will carry the compass to 10 degrees west, the return swing to $3\frac{1}{3}$ degrees east, then $1\frac{1}{9}$ degrees west, and so on, until it comes to rest. Fig. 19 shows graphically the damping characteristics of the gyro-compass.

Period of Oscillation

The natural period of the compass, i.e., the time it takes to perform a complete oscillation (from A to B in Fig. 18), is 85 minutes. The period of oscillation is governed by two factors:

1. The angular momentum of the gyro (the product of the weight, speed and square of radius of gyration) and
2. The torque about the horizontal axis supplied by the action of the mercury ballistic. (This, in turn, is governed by the free surface of the mercury in the containers and the distance of the containers from the horizontal axis.)

If the weight or the speed of the gyro is increased, the period of oscillation will be longer. If the free surface of the mercury is increased, the period of oscillation will be shorter.

The compass is given a period of oscillation which is many times the period of the rolling movement of the ship; otherwise the compass might show some deviation before such movements could reverse and cancel each other.

Compensating Weights

Unsymmetrical distribution of weight is another potential source of disturbance, when the compass is swinging, which must be neutralized. When the unsymmetrical weight shown in Fig. 20

is swung in the plane A–B, centrifugal stresses act upon it in such a way as to cause all of its particles to place themselves as far as possible from the axis of swing. This causes a tendency to turn, as indicated by the arrows. The same effect may be observed when a watch is swung back and forth on its chain through a small arc. In the gyro-compass this effect is avoided by the use of compensating weights which permit of a symmetrical distribution of weight about the vertical axis.

The foregoing explanation applies particularly to a compass at the equator and in a vessel which is not under way. At points other than the equator and on board ships which are moving over the surface of the earth, certain factors are introduced that would result in errors if they were not compensated or corrected in the design of the compass.

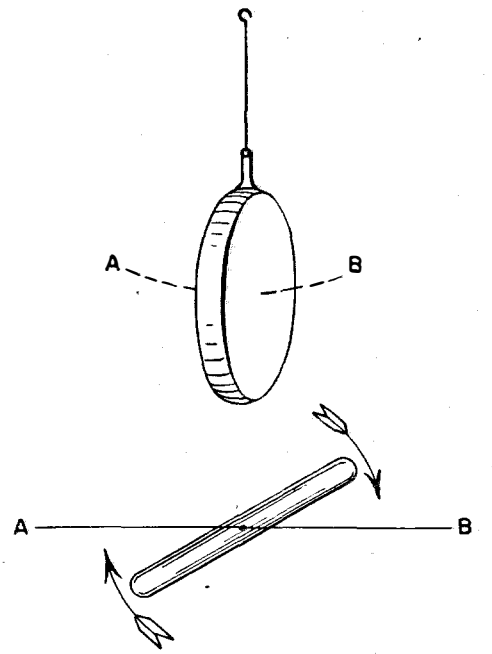

Fig. 20—Effect of unsymmetrical distribution of weight.

Latitude Correction

Latitude correction is necessary because of the eccentric connection employed to damp the oscillations of the compass. The correction is made by means of a latitude adjustment scale, no special knowledge of the problem being required in order to make the correction. A complete explanation of the reasons for the latitude error, however, will be welcomed by the student because it involves a general discussion of compass behavior and will give the student an opportunity to find out some of the whys and wherefores that might otherwise escape attention.

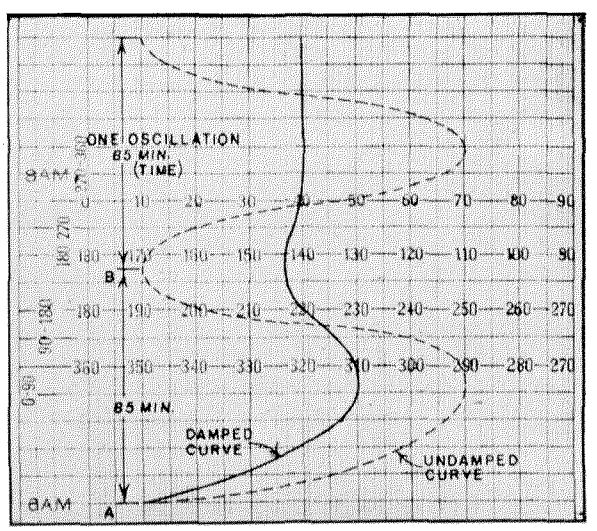

Fig. 19—Gyro-Compass damping curve as charted by a course recorder operated from the master compass. A shows settling characteristics when compass is set 30° away from the meridian. B shows undamped oscillations. Chart is read from bottom up.

At the equator, where only the horizontal component of the earth's rotation affects the gyro, the axle of the compass is horizontal and parallel to the earth's axis. At the equator, therefore, as soon as the compass has settled on the meridian, the ballistic will be at rest and the compass may be considered as a true gyroscope.

If we move the compass to a point to the north or south of the equator, however, it will be affected by the vertical as well as the horizontal component of the earth's rotation. At a point north of the equator, for instance, the north end of the gyro axle tends to turn toward the east and rise, as the earth rotates out from beneath it. This was illustrated in Fig. 13. It is apparent that the north end of the gyro axle must be precessed continuously in a westerly direction toward the meridian as fast as it is being displaced by the vertical component of the earth's rotation. The force necessary to do this is obtained automatically by the simultaneous tilt of

the mercury ballistic which permits an accumulation of mercury in the lower or south containers so as to precess the axle continuously toward the meridian.

If it were not for the offset connection of the mercury ballistic, the gyro axle would eventually settle, theoretically, exactly on the meridian, with the end away from the equator tilted up just enough so that there would be sufficient excess mercury in the containers toward the equator to precess the gyro toward the meridian at the same rate as it is being carried away by the vertical component of the earth's rotation.

In order to damp the oscillation, however, the mercury ballistic connection is offset to the east of the centerline of the compass. This produces a counter-clockwise torque about the vertical axis, causing the north end of the axle to precess continuously *down*. The axle, therefore, settles to the east of the meridian (in north latitudes) at a point where the downward precession of the axle due to torque about the vertical axis is exactly balanced by the horizontal component of the earth's rotation tending to tilt the north end up.

As we move the compass further north, the north end of the gyro axle turns to the east faster and rises faster, and for this reason the compass must be precessed faster toward the meridian. The downward pressure on the south end of the axle is correspondingly greater, the torque about the vertical axis is greater, the north end precesses down faster, and the gyro consequently settles further to the east.

South of the equator the effect of the earth's rotation on the gyro is just the opposite: here the *south* end of the axle tends to rise and turn to the east; the excess mercury is in the north containers, precessing the *north* end to the east toward the meridian. Torque about the vertical axis is reversed, so that the south end is being precessed down at the same rate as the horizontal component of the earth's rotation is tilting it up. Therefore the north end of the axle lies slightly to the *west* of the meridian.

The small angle at which the gyro axle settles from the meridian varies with the latitude, and for this reason a correction must be introduced to compensate for this natural error at any latitude where the compass may be expected to be used.

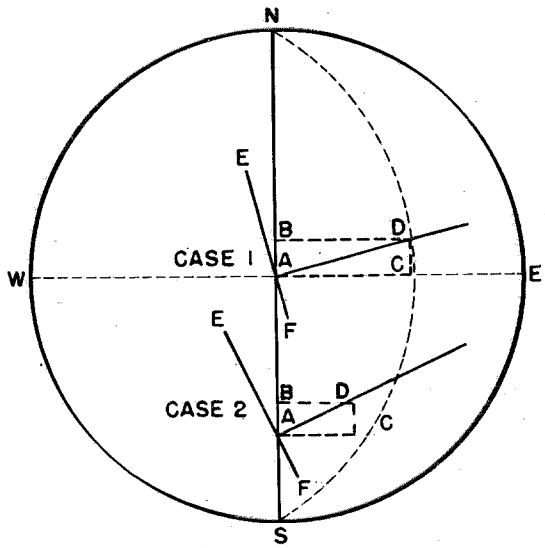

Fig. 21—Diagram illustrating the effect of ship's speed and latitude on the gyro-compass.

From zero at the equator, this error increases to 2.9 degrees at 60 degrees north or south latitude; it is easterly in north latitudes and westerly in south latitudes.

No attempt is made in the gyro-compass to compel the axle to seek a resting place at variance with its natural settling position. Instead, a latitude adjustment is provided which moves the lubber ring the necessary amount to cause the compass indication to be true. Since the transmitter is mounted on the lubber ring a correctional movement applied to the lubber ring also transmits the corrected reading to the repeater compasses, course recorder and gyro-pilot for automatic steering. This is the "Latitude Correction," and is made independently of the speed and course correction, described below.

Ship's Speed and Course

It has been shown that the relatively slow angular movement of the earth's rotation (only one revolution in 24 hours) provides the motive force for the north-seeking precessional movement of the compass. When a vessel is traveling over the earth's surface, however, and therefore about the earth's center, the vessel's movement is compounded with that of the earth. If the direction of travel is east or west, the vessel's

motion only adds to or subtracts from the earth's motion, and the effect on the indication of the compass is negligible.

When the vessel is traveling north or south, however, the effect is a maximum because the ship's speed produces a resultant which is not parallel to the plane of the earth's rotation. The effect on compass indication is proportional to the ship's speed and course and is explained as follows:

In Fig. 21 (Case 1) the line AB represents the movement of a vessel steaming due north for a given period of time. Line AC represents the movement of the earth in the same period of time. The actual path taken by the vessel relative to the earth's rotation is AD. AC is the normal plane of rotation. AD is the actual plane of rotation due to combined movement of earth and ship. As far as the gyro-compass is concerned, therefore, a new meridian EF is produced which is at right angles to the line AD. At higher latitudes the movement of the earth is relatively smaller, the new, apparent meridian is displaced farther from the actual meridian, and a larger correction is required. Thus the amount of the correction is dependent upon the latitude of the ship. (Case 2 of Fig. 21.)

The displacement of the gyro axle is to the west for northerly courses and to the east for southerly courses, and the angular difference between the actual meridian and this new apparent meridian depends upon:

1. The ship's speed.
2. The ship's course, as it is only the northerly or southerly components of the course which are to be taken into account, and—
3. The latitude, as the higher the latitude the smaller the earth's surface speed and therefore the greater the effect of the ship's speed.

The compass is provided with a corrector mechanism which automatically applies the correction due to course and speed at any given latitude by moving the lubber line the required amount to compensate for this error. The compass indicates the ship's true heading on all courses, and no corrections have to be applied by reference to tables.

Plastics—A Brief Review of Their Physical and Electrical Properties

By A. J. WARNER, B.Sc. (London)
Federal Telephone and Radio Corporation, Newark, New Jersey

WITHOUT the development of the vacuum tube, modern achievements in the radio art would have been impossible. Not so spectacular, but nonetheless very important, has been the contribution made by the chemist and chemical engineer in the development and production of plastic materials. All electronic equipment currently being manufactured or in course of development will be found to incorporate in one form or another plastic materials as vital, constituent parts.

Apart from the electronic industry, other important electrical industries, such as power distribution, are also very large users of plastics, although in this case they usually differ in type from those necessary for high frequency work. It would seem advantageous, therefore, to discuss the various types of plastic materials available, paying particular attention to their electrical properties; and, also, to examine their physical limitation since there is no "all purpose" plastic available for the variety of conditions encountered. In all cases, a compromise must be made between the electrical properties desired and the mechanical limitations inherent in the material.

The last few years have seen a tremendous increase in the number of types of plastic materials, as is reflected in the ever growing literature and patent art on the subject, but much of the knowledge is specialized and not easily available to the general technician who, under stress of modern times, has neither the facilities nor the time to search out the pertinent data, but must rely on summaries prepared by others to guide him in the choice of suitable materials. It is hoped that this article may partially fill such a need.

In a previous article,[1] we examined the question of the two most important properties of plastics for use in electrical systems—namely, the dielectric constant and the power factor. But there are other properties that may assume greater importance for specific applications: dielectric strength, volume resistivity, arc resistance, etc. Among the chief physical properties that require examination are heat distortion, cold temperature resistance, flexural strength, tensile properties, aging characteristics, moisture absorption, flammability, coefficient of thermal expansion, etc.

Owing to the comparative newness of the plastic art, it was not until recently that any concerted effort was made to standardize on methods of test or to evaluate at all fully the many varieties of plastics. Such work is now being conducted by various societies and associations, and some material has already been published. Reference may be made to the following excellent publications:

1. A.S.T.M. Standards on Plastics, October, 1943.
2. Technical Data on Plastic Materials, published by the Plastics Materials Manufacturers Assn.
3. 1944 Plastics Catalog.

The Society of the Plastics Industry also is currently working on an engineering classification of plastic materials through a specially organized engineering committee. It is expected that the conclusions of this committee, when published, will be of great value to designers and engineers.

The term "plastics" is a very loose one and scientifically includes any material that can be deformed under mechanical stress without losing its coherence and is able to keep the new form given it. The newness of the art, however, is such that no clear-cut definition of a plastic material has yet received universal recognition; the definition just given will cover materials such as clay, bitumen, glass, etc., which many persons will not consider plastic materials. For purposes of discussion, however, we will concern ourselves only with materials either synthesized in the laboratory or modified from naturally occurring materials which are capable of being fabricated into useful shapes by the use of heat and pressure.

Over a long period of time, it has become customary to divide such plastic materials into two main groups, the so-called heat infusible or thermosetting materials, and the heat fusible or thermoplastic materials. This arbitrary separation is nowadays becoming less important, since we are learning how to make materials which can be changed from one form to the other by suitable processing. Of these two groups, the thermoplastic type of material is becoming more and more important in electrical communication systems. Not only does this group possess very superior electrical properties, but it does not suffer from limitations of fabricated parts as regard size and intricacies of shape.

Thermosetting Resins

1. Phenol Formaldehyde Resins

One of the best known examples of a thermosetting material, and one which, moreover, has found considerable use, is "Bakelite," or phenol formaldehyde resin. The preparation of an insoluble resinous body formed by the interaction of phenols and aldehydes dates back to the middle of the nineteenth century, when A. Baeyer [2] in 1872 announced that such a reaction was general between phenols and aldehydes. He noticed that benzaldehyde united with pyrogallic acid on heating in a manner similar to that with phthalic acid, and a colorless resinous product resulted.

The chemistry of the phenol formaldehyde resins in the various stages of their formation is very complex and is still the subject of much discussion and argument, but, before the infusible, final product is reached, the material goes through several thermoplastic stages which enable forming under pressure to take place. The majority of resins of this type used today in moulding compounds are derived from phenols, cresols, or xylenols and formaldehyde. Cured phenolic resins develop great strength and hardness and are relatively resistant to heat, water, organic solvents, and mild alkalies. They are rather hard and brittle, however, and in order to obtain satisfactory impact strengths, particularly at low temperatures, special care must be taken to use carefully chosen fillers to achieve the desired properties. In general, the phenolic resins have a greater coefficient of expansion than metals, and thus tend to shrink around metal inserts after moulding and to hold them in position tightly.

One of the drawbacks to the use of thermosetting materials in general has been the difficulty in moulding large pieces or thick sections because of the difference of cure found between the outside of the piece in contact with the hot mould and the center of the piece. This difference of cure results in shrinkage of the part after moulding, as well as considerable strain inside the part itself, and generally unsatisfactory physical properties. Development of Megatherm electronic pre-heating apparatus, however, has provided facilities for overcoming this difficulty to a great degree, and has given a great stimulus to the extended use of these materials.

In view of the wide diversity of properties that can be obtained by the use of suitable compounds and fillers, phenolic resins are usually classified into groups depending on the property for which they are principally developed. An examination of some of the typical groups of phenolic resins will, therefore, be interesting.

A. *Cast Phenolic Resins*

Cast phenolic resins are prepared by pouring a liquid processed phenol formaldehyde resin into a straight-draw open mould or split mould. These moulds are usually made from lead, and the fillings are placed in ovens to complete the cure which may involve heating for periods ranging from a few hours to six days. The castings on removal from the mould are then machined into finished articles. Such cast phenolic resins are usually used for decorative purposes, although certain compounds have been produced with reasonably good dielectric properties. Cast phenolic resins are known under the trade names of Bakelite, Catalin, Prystal, Durez, Marblet, Opalon, Gemstone, Baker Resin, etc. The average properties of this type material are indicated in Table I.

B. *General Purpose Cellulose Filled Phenolic Resins*

General purpose cellulose filled phenolic resins are prepared by adding fillers such as wood flour, cotton flock, macerated fabric, etc., to the resin compound and thoroughly kneading on revolving heated rollers. This rolling homogenizes the ma-

terial and also advances the resin toward a more cured condition. The material, as it leaves the blending rolls, is in the form of rough sheets which are subsequently cooled and ground to the correct particle size, ready for moulding. Such materials can be moulded with conventional type moulding presses. Cellulose filled phenolic resins are sold under the trade name of Bakelite, Durez, Durite, Makalot, Heresite, Resinox, Indur, etc. Table II gives average properties of this type of material.

C. *High Frequency Mineral Filled Phenolic Resins*

Owing to the relatively poor electrical performance of the usual fabric filled phenolic resins, attempts have been made to improve the electrical properties of phenolic resins by the inclusion of mineral fillers. Such materials are often referred to in the trade as "mica." They must be handled carefully in processing owing to the difficulty of getting a completely uniform cure. Their specific gravity is relatively high, ranging from 1.80 to 1.92, but their water absorption is lower than most other phenolic types. The physical and electrical properties of the best mineral filled phenolic resins available are shown in Table III.

2. Urea Formaldehyde Resin

A further type of thermosetting resin is prepared by the condensation of urea with formaldehyde under controlled conditions of time, temperature, and acidity. The intermediate stages of the resins are water soluble and, under certain circumstances, alcohol soluble. As a water soluble resin, the intermediate compounds are often used to treat paper, cloth, and wood veneer for laminating. Mixed with finely divided alpha cellulose, the urea resins may be formed into moulding powders. One of the chief advantages of the urea resins over the phenol formaldehyde type materials is the fact that they can be prepared in pastel shades, whereas the phenolics are usually available only in dark colors. The materials are sold under the trade name of Bakelite, Beetle, Plaskon, Urea, and Uformite. Urea formaldehyde moulding compound materials have a hard surface finish and are, therefore, particularly useful for the manufacture of articles subject to considerable handling and wear. Their use, however, is limited to conditions where they will not be exposed to temperatures above 180°F. Electrically the urea formaldehyde resins are somewhat superior to the common phenolic types since they possess relatively high dielectric strength, high arc resistance, no tendency to track after arcing, and a relatively low power factor. They are not, however, so good electrically as the mineral filled phenolics. The electrical properties are little affected by exposure to high humidity. (See Tables IV and V.)

3. Melamine Formaldehyde Resin

One of the newer types of thermosetting materials is the melamine formaldehyde type of

TABLE I

Cast Phenolic Resins, Average Properties

Impact strength ft. lbs. per in.[3]	.48 to .60
Tensile strength psi[4]	8500 to 10,000
Modulus of elasticity in tension psi[5]	3.75 to 4.5×10^5
Rockwell hardness[6]	M65 to M80
Specific gravity[7]	1.30 to 1.32
Heat distortion point °F.[8]	113 to 176
Water absorption after 24 hour immersion %[9]	.5
Dielectric strength short time, 1/8", v/m[10]	350 to 430
Volume resistivity ohms-cm.[11]	1 to 7×10^{12}
Dielectric constant[12]	
60 cycles	7.0 to 8.0
1000 cycles	6.0 to 7.0
1 megacycle	5.5 to 6.0
Power factor[13]	
60 cycles	.10 to .15
1000 cycles	.03 to .04
1 megacycle	.03 to .05
Arc resistance[14]	200 to 250 seconds

TABLE II

Cellulose Filled Phenolic Resins, General Purpose—Average Properties

Specific gravity	1.30 to 1.47
Impact strength ft. lbs. per in.	.24 to .32
Tensile strength psi	7000 to 8000
Modulus of elasticity in flexure psi	8 to 11×10^5
Rockwell hardness	M115 to M120
Coefficient of thermal expansion °C.	3.0 to 3.5×10^5
Heat distortion point °F.	275 to 284
Water absorption after 24 hour immersion %	.40 to .75
Dielectric strength short time, 1/8", v/m	300 to 350
Volume resistivity ohms-cm.	1 to 5×10^{11}
Dielectric constant	
60 cycles	7.5 to 15
1000 cycles	6.0 to 9.5
1 megacycle	5.0 to 6.8
Power factor	
60 cycles	.25 to .50
1000 cycles	.07 to .25
1 megacycle	.04 to .06
Arc resistance	Poor

material. Its excellent shock and heat resistant properties combined with high arc resistance and lack of tendency to track after arcing has made it of great value in many diverse applications, including aircraft ignition parts, moulded circuit breakers, terminal blocks, etc. The resins only became commercially available in the United States around 1939, although melamine itself was discovered by Liebig as early as 1834. Melamine resins, in combination with alpha cellulose fillers, are similar in appearance to the urea formaldehydes, but they show greater resistance to acids and alkalies and are less affected by boiling water and are considerably harder than the corresponding ureas. With cotton or rag filler, the melamine resins yield products of high flexural strength, low water absorption, good arc resistance, and general inertness. Within a temperature range of −70°F. to +210°F., their physical properties are not appreciably changed. In solution form, the resins can be used to impregnate fabrics and paper for use as laminates. Heavy duty telephone hand sets manufactured for the United States Navy and Marine Corps are moulded of an alpha cellulose filled melamine plastic. The resins are sold under the trade name of Melmac, Plaskon, and Catalin. (See Tables VI and VII.)

4. Phenol Furfural Resins

With the increasing availability of chemical products derived from such cheap sources of raw material as corn stalks, it was natural that an attempt should be made to utilize these products in resin manufacture. One of the outstanding developments reaching the commercial stage has been the use of furfural as a replacement for formaldehyde, which was becoming a very serious bottleneck in the manufacture of thermosetting types of resins owing to the large requirements of the armed services for the starting material used in the manufacture of formaldehyde. Apart from its ability to react with phenol, furfural itself plays an important rôle in the formation of a network of cross-linked molecular chains in the resin itself, and thus provides products capable of a high degree of completeness in cure and possessing high strength, good impact resistance, and generally good chemical resistance. By the choice of the proper fillers, mouldings made from these materials can meet a wide variety of mechanical and electrical requirements. They also possess the advantage of being non-inflammable and grip metal inserts tightly. Furfural phenol resins, in a considerable number of grades, are sold under the trade name of Durite. (See Tables VIII and IX.)

TABLE III

High Frequency Mineral Filled Phenolic Resins

Impact strength ft. lbs. per in.	.4 to .5
Tensile strength psi	5500 to 7000
Modulus of elasticity in flexure	30 to 40×10^5
Rockwell hardness	M105 to M110
Heat distortion point °F.	212 to 311
Water absorption after 24-hour immersion %	0.3 to 0.7
Dielectric strength short time, 1/8", v/m	400 to 450
Volume resistivity ohms-cm.	Greater than 10^{-4}
Dielectric constant	
60 cycles	5 to 5.2
1000 cycles	4.9 to 5.1
1 megacycle	4.8 to 5.0
Power factor	
60 cycles	.01 to .02
1000 cycles	.009 to .015
1 megacycle	.005 to .009
Arc resistance	Low

TABLE IV

General Purpose Urea Resins

Impact strength ft. lbs. per in.	.24 to .36
Tensile strength psi	6000 to 13,000
Specific gravity	1.45 to 1.55
Heat distortion point °F.	260 to 280
Water absorption after 24-hour immersion %	.75 to 3.0
Dielectric strength short time, 1/8", v/m	300 to 400
Volume resistivity ohms-cm.	10^6 to 10^{13}
Dielectric constant	
60 cycles	7.0 to 9.5
1000 cycles	6.4 to 9.0
1 megacycle	6.3 to 7.5
Power factor	
60 cycles	.035 to .10
1000 cycles	0.035 to 0.055
1 megacycle	0.027 to 0.04

TABLE V

Alpha Cellulose Filled Urea Formaldehyde Resins

Impact strength ft. lbs. per in.	.24 to .36
Tensile strength psi	6000 to 13,000
Modulus of elasticity in flexure	12 to 15×10^5
Rockwell hardness	M118 to M122
Specific gravity	1.45 to 1.55
Heat distortion point °F.	260 to 280
Water absorption after 24-hour immersion %	.75 to 3.0

The electrical properties of the alpha cellulose filled resins are very similar to those for the general purpose material (Table IV).

TABLE VI

ALPHA CELLULOSE FILLED MELAMINE RESINS

Impact strength ft. lbs. per in.	.26 to .28
Tensile strength psi	About 5000
Specific gravity	1.49
Heat distortion point °F.	385
Water absorption after 24-hour immersion %	1.0 to 1.7
Dielectric strength short time, 1/8", v/m	340
Dielectric constant	
60 cycles	7.5 to 8.3
1 megacycle	6.7 to 7.3
Power factor	
60 cycles	0.025 to 0.050
1 megacycle	0.028 to 0.029
Arc resistance	125 seconds

TABLE VII

MINERAL FILLED MELAMINE RESINS

Impact strength ft. lbs. per in.	.28 to .40
Tensile strength psi	5500 to 7000
Modulus of elasticity in flexure	$.6\times10^6$
Rockwell hardness	M112
Specific gravity	1.70 to 2.00
Heat distortion point °F.	266
Water absorption after 24-hour immersion %	0.08 to 0.14
Dielectric strength short time, 1/8", v/m	390
Volume resistivity ohms-cm.	2.4×10^{11}
Dielectric constant	
60 cycles	6.4 to 9.9
1 megacycle	6.7
Power factor	
60 cycles	0.07 to 0.17
1 megacycle	0.041
Arc resistance	120 to 140 seconds

The arc resistance of both the fabric and mineral filled furfural phenol materials is not very good, being no better than that of a standard phenol formaldehyde type of material.

5. CASEIN PLASTICS

The development of plastic materials from casein, a protein material derived from milk, is due to the work of two German chemists, Spitteler and Krische, who around 1900 found that the immersion of casein in formaldehyde gave a hard material which, after drying, could be machined and turned into a variety of shapes. The trade name, Galalith, was given to this material, and this name is still in common usage throughout the world. Owing to the development of other types of plastics and the fact that casein plastics are very susceptible to changes in atmospheric conditions, etc., the use of casein has been limited and finds its chief outlet in the manufacture of small decorative articles such as buttons, buckles, etc. Casein plastics have excellent machining qualities and are easily polished. By reason of its chemical constitution, casein plastic is non-inflammable and can be manufactured in a wide range of colors; it is sold in the U. S. A. under the trade name of Ameroid and Galorn. Because of its limited applications, there are not much data on it available. A few of its general properties are given in Table X.

Electrically, this material is very poor and coupled with its high degree of hygroscopicity, it is not usually considered for electrical application.

TABLE VIII

FABRIC FILLED PHENOL FURFURAL RESINS

Impact strength ft. lbs. per in.	1.20 to 4.60
Tensile strength psi	5500 to 8000
Modulus of elasticity in tension psi	$7 \text{ to } 12\times10^5$
Rockwell hardness	B65 to B75
Specific gravity	1.3 to 1.4
Heat distortion point °F.	About 270
Water absorption after 24-hour immersion %	0.8 to 1.4
Dielectric strength short time, 1/8", v/m	150 to 450
Volume resistivity ohms-cm.	10^9 to 10^{11}
Dielectric constant	
60 cycles	5 to 10
1000 cycles	5 to 9
1 megacycle	4.5 to 8
Power factor	
60 cycles	0.06 to 0.30
1000 cycles	0.06 to 0.17
1 megacycle	0.03 to 0.1

TABLE IX

MINERAL FILLED PHENOL FURFURAL RESINS

Impact strength ft. lbs. per in.	0.26 to 1.0
Tensile strength psi	4000 to 8000
Modulus of elasticity psi	$10 \text{ to } 45\times10^5$
Rockwell hardness	B55 to B65
Specific gravity	1.6 to 2.0
Heat distortion point °F.	275 to 295
Water absorption after 24-hour immersion %	0.01 to 0.15
Dielectric strength short time, 1/8", v/m	250 to 450
Volume resistivity ohms-cm.	10^9 to 10^{11}
Dielectric constant	
60 cycles	5 to 15
1000 cycles	5 to 14
1 megacycle	4.5 to 12
Power factor	
60 cycles	0.02 to 0.28
1000 cycles	0.02 to 0.15
1 megacycle	0.01 to 0.1

TABLE X

CASEIN PLASTICS, GENERAL PROPERTIES

Impact strength ft. lbs. per in.	1.0
Tensile strength psi	10,000
Modulus of elasticity psi	5.1 to 5.7 × 10^5
Elongation %	2.5
Specific gravity	1.35
Softening point °F.	200
Water absorption after 24-hour immersion %	7 to 14
Dielectric strength short time, 1/8", v/m	400 to 700
Dielectric constant 1 megacycle	6.2 to 6.8
Power factor 1 megacycle	0.052

6. LAMINATED PLASTICS

A logical development of the phenolic plastics, particularly where large sheets of varying thicknesses are required, was the impregnation of fibrous sheet materials with the uncured resin and then the consolidation of these sheets into hard products by pressure and heat. The fibrous sheets principally employed for these purposes are paper, cotton fabric, canvas duck, and asbestos. More recently, glass has been used as the filling material. For purposes of standardization, laminated sheets are manufactured under the following grades:

Grade X. This is a general purpose paper base laminated material for use where electrical properties are not so important as physical properties. In thinner sections, the material can be punched cold and to greater thicknesses when heated. The material machines and saws readily.

Grade P. This is a general purpose paper base laminated material especially developed for punching operations, being more flexible but not quite so strong as Grade X.

Grade XX. This is a paper base laminated material for electrical applications requiring low moisture absorption. It has good machineability.

Grade XXP. This is a paper base laminated material similar to Grade XX but, being somewhat more flexible, it is suitable for hot punching.

Grade XXX. This paper base laminated material has extremely low moisture absorption and high dielectric strength, and was designed to be suitable for radio frequency applications at high humidities.

Grade XXXP. This is similar to Grade XXX but has somewhat lower dielectric losses and is more suitable for hot punching.

Grade C. This is a canvas base laminate designed for structural applications where high tensile and transverse strength are required. It is suitable for the manufacture of gears requiring high impact.

Grade CE. This is a fabric base laminated material similar to Grade C designed for electrical applications where greater toughness than Grade XX is required or for mechanical applications where greater resistance to moisture is required than for Grade C.

Grade L. This is a linen base fine-weave laminate suitable for small gears and other fine machining applications. It is not so tough as Grade C. Electrically it is poor and should not be used, except for very low voltage.

TABLE XI

LAMINATED PHENOLIC MATERIALS

Name of Material	Dielectric Strength Minimum	Dielectric Constant 1 Mc.	Power Factor 1 Mc.	Compressive Strength Psi Average	Flexural Strength Psi Average	Tensile Strength Psi Average	Water Absorption Maximum
X	360	—	—	35,000	21,000	12,500	3.3
P	360	—	—	33,000	15,000	8,000	2.8
XX	360	5.5	0.045	34,000	16,000	8,000	1.3
XXP	360	5.5	0.045	25,000	16,000	8,000	1.3
XXX	360	5.2	0.035	32,000	15,000	7,000	.85
XXXP	360	5.2	0.030	25,000	15,000	7,000	.85
C	150	7.0	0.10	38,000	20,000	9,500	2.5
CE	290	6.0	0.065	36,000	17,000	8,000	1.4
L	150	7.	0.10	35,000	20,000	9,000	1.6
LE	290	5.5	0.055	37,000	19,000	8,500	1.25
A	160	—	—	36,000	16,000	8,000	.95
AA	50	—	—	38,000	20,000	10,000	.95

Grade LE. This is a fine-weave fabric base laminate similar to Grade L designed for electrical applications requiring toughness better than Grade XX. It is exceptionally good as regards moisture resistance.

Grade A. This is an asbestos paper base laminated material having high resistance to flame and being slightly more resistant to heat than other laminated grades, but it is suitable only for low voltage applications.

Grade AA. This is similar to Grade A except that it is stronger and tougher.

For purposes of comparison, Table XI shows the physical and electrical properties of the various laminated phenolic materials.

7. Aniline Formaldehyde

One of the more recent developments in the field of aldehydic thermosetting resins is the aniline formaldehyde resin, known in the trade as Dilectene or Cibanite. Because of the improved dielectric properties, coupled with resistance to water and other chemicals, of this type material over many types of phenol formaldehyde compounds, the resins are finding increasing use, particularly in the fields of military communication. The pure resin is thermoplastic in nature, but not to the same extent as the true thermoplastics. It can be fabricated with standard equipment provided care is taken to see that the tools and material being fabricated are kept cooled. It can be tapped, turned, drilled, machined, sawed, polished, and threaded quite readily, provided the cutting tools are sharp and the operating speed is kept high. It is not particularly affected by atmospheric conditions or ultra violet light and is reasonably tough so that it is finding its place as a replacement for hard rubber or ebonite and ceramic and bonded mica combinations. Table XII gives data on aniline formaldehyde resins.

TABLE XII

Aniline Formaldehyde

Coefficient of thermal expansion	3×10^{-5}
Tensile strength psi	10,500
Rockwell hardness	M100 to M125
Specific gravity	1.21
Heat distortion point °F.	210
Water absorption after 24-hour immersion %	0.08
Dielectric strength short time, 1/8", v/m	640
Volume resistivity ohms-cm.	10^{12} to 10^{13}
Dielectric constant	
60 cycles	3.7 to 3.8
1000 cycles	3.7
1 megacycle	3.5 to 3.6
Power factor	
60 cycles	0.002
1000 cycles	0.004
1 megacycle	0.006 to 0.008

One of the chief drawbacks of aniline formaldehyde resin [5] is its sensitivity to moderately high temperatures. Although attacked by strong acids, it is unaffected by alkalies and reasonably resistant to the common organic solvents; alcohol and aromatic hydrocarbons particularly have no affect on the material. It has found considerable use in aircraft radio work for the manufacture of coil forms, vacuum tube sockets, antenna housings, etc. Because of its good insulation resistance, which is maintained under the influence of moisture and atmospheric conditions, the material has been found very useful for terminal boards, mounting strips, etc.

Thermoplastic Resins

Perhaps the greatest strides in recent years in the plastic field have been made in the domain of the so-called thermoplastic type materials. The delay in the commercial preparation of these materials was chiefly due to the fact that successful synthesis on an economic scale was dependent on engineering methods and knowledge of materials which were not known before the twentieth century. Thus, we find that monomeric styrene, from which is manufactured that very important thermoplastic material, polystyrene, was actually discovered as long ago as 1829, but it was not until 1935 that the commercial manufacture of this material on any appreciable scale was commenced in the U. S. A. So rapid has been the advance in this field, however, that at the present moment it is true to say that the production of styrene far exceeds that of any other single type plastic material or, for that matter, of any synthetic organic chemical. It is also in the field of thermoplastic materials that one finds the most intense development, and scarcely a month goes by without the announcement of some new material having "improved" properties over that of any existing type. It is our belief that future

developments in the plastics field will tend more and more toward the use of thermoplastic materials and away from the conventional thermosetting materials. We will see as we examine the various types of thermoplastic materials that a very wide range of physical properties can already be obtained and, in many cases, properties that are unobtainable with the older thermosetting type compounds. Among typical thermoplastic materials currently used in large quantities are polystyrene, polymethyl methacrylate, polyvinyl chloride, polyethylene, polybutene, nylon, etc. The thermoplastic materials owe their popularity and wide use to the variety of physical and electrical properties obtained therewith combined with ease of fabrication by automatic moulding machines, generally good dimensional stability, good aging characteristics, availability in clear, natural, or pastel shades; and, with certain special types, vastly superior electrical characteristics over a wide temperature and frequency range. Thermoplastic materials themselves cover a variety of physical properties from sticky materials through highly elastic bodies to tough, solid compounds; also, by virtue of their chemical constitution or special formulation, they can be made flame resistant where desired.

1. Cellulose Derivatives

Plastic materials derived from cellulose have always been a potentially fruitful source of investigation due in large measure to the considerable cheap raw material supplies available in cotton, wood, etc., and it is not surprising, therefore, that many plastic materials have been made from this starting commodity. The chief ones being commercially exploited and with which we shall concern ourselves here are (A) cellulose nitrate, (B) cellulose acetate, (C) cellulose acetate butyrate, and (D) ethyl cellulose.

A. *Cellulose Nitrate*

This inorganic ester of cellulose is the oldest plastic cellulose derivative. The discovery of nitrocellulose is usually credited to Schönbein who, in 1845, nitrated cellulose with a mixture of nitric and sulphuric acids. The commercial exploitation of this material as a plastic substance is chiefly due to the work of Hyatt, who plasticized the material with camphor and alcohol to obtain a flexible material. Because of the highly inflammable nature of this plastic and its tendency to yellow and brittle with age, the material has limited applications and, in particular, is not usually employed in electrical systems. At room temperature, it is the toughest thermoplastic material known, and it is also very resistant to moisture. One of the great disadvantages to its use is that it is not possible to mould the material by conventional methods. Cellulose nitrate plastics are sold under the trade names of Celluloid, Pyralin, Nitron, and Nixonoid. Their properties are listed in Table XIII.

TABLE XIII

Cellulose Nitrate Plastics

Impact strength ft. lbs. per in.	4.5 to 5.5
Tensile strength psi	6000 to 7500
Coefficient of thermal expansion	12 to 16×10^{15}
Rockwell hardness	M25 to M30
Specific gravity	1.39 to 1.45
Heat distortion point °F.	140
Water absorption after 24-hour immersion %	1.5 to 2
Dielectric strength short time, 1/8″, v/m	300 to 600
Volume resistivity ohms-cm.	10 to 15×10^{10}
Dielectric constant	
60 cycles	6.7 to 7.3
1 megacycle	6.2
Power factor	
60 cycles	0.08 to 0.12
1 megacycle	0.07 to 0.10

B. Cellulose Acetate

Cellulose acetate is one of the most widely known of plastic materials, being used in large quantities as transparent sheeting, photographic films, wrapping material, etc. It was first discovered by Schutzenberger[15] in 1865, and the process was considerably improved by Franchimont[16] in 1879. Cellulose acetate itself is somewhat brittle, and in order to prepare suitable foils and films that will withstand the flexing necessary, plasticizers must be added to the base plastic. Naturally, the final physical and electrical properties will depend greatly on the nature and amount of the plasticizer employed. Cellulose acetate is sold under the trade names Plastacele, Tenite I, Nixonite, Lumarith, etc. The average properties of cellulose acetate sheet are given in Table XIV.

TABLE XIV

CELLULOSE ACETATE SHEET, AVERAGE PROPERTIES

Coefficient of thermal expansion	8 to 16×10^{-5}
Tensile strength psi	4000 to 14,000
Modulus of elasticity psi	1 to 3.5×10^{5}
Elongation %	20 to 55
Specific gravity	1.36
Heat distortion point °F.	100 to 190
Water absorption after 24-hour immersion %	2 to 4
Dielectric strength short time, 1/8", v/m	290 to 325
Dielectric constant	
60 cycles	4.9
1 megacycle	3.7
Power factor	
60 cycles	0.016
1 megacycle	0.044

C. *Cellulose Acetate Butyrate*

Recent developments in the cellulose field include the preparation of a mixed ester of cellulose, cellulose acetobutyrate, giving a new thermoplastic material whose properties in many respects are superior to those of cellulose acetate itself; it has found considerable and increasing use in the last few years. Cellulose acetobutyrate is made by the esterification of alpha cellulose with a mixture of acetic and butyric acids and anhydrides in the presence of a catalyst. The properties of the resultant material can be varied by altering the relative proportions of the acids and anhydrides used. In general, the mixed ester has a greater solubility in a wider range of solvents and can be plasticized more readily than the straight ester. Moisture absorption and weather resistance characteristics are improved in the mixed ester composition, and lacquers made with these compositions have greater adhesion than those prepared with cellulose acetate. A considerable amount of cellulose acetobutyrate foil is used today as a primary insulation on certain wires and cables, thus permitting an important reduction in the diameter of wires with corresponding saving in weight and space. The material is marketed under the trade name of Tenite II. Because of the wide range of materials available, it is not possible to list all the properties of the varying grades, and we shall content ourselves with listing two typical examples (Table XV). In general, however, the electrical properties of these materials are substantially the same.

D. *Ethyl Cellulose*

A relative newcomer in the plastics field and one which undoubtedly will find wide application is the ethyl ether of cellulose, ethyl cellulose. The history of cellulose ethers goes back to 1905 with the work of Suida.[17] Technical development dates from 1912 with the filing of patents simultaneously by Leuchs,[18] Dreyfus,[19] and Lilienfeld.[20] It was not until 1941 and 1942, however, that any large scale utilization of the material was made. Like other thermoplastic materials, ethyl cellulose can be especially formulated to meet specific requirements, the chief improvement being in low temperature flexibility properties. Although some plasticization of the material is carried out, the pure resin itself possesses remarkable flexibility. Three main types of ethyl cellulose plastics are available (see Table XVI).

From Table XVI, it is evident that the electrical properties of this material, sold under the trade names Ethocel and Hercules E.C., are superior to the vinyl chlorides or acrylate resins, but they by no means approach polystyrene or polyethylene. The low temperature flexibility is an outstanding property.

2. VINYL RESINS

A very important group of thermoplastic materials are the vinyl compounds of which vinyl

TABLE XV

CELLULOSE ACETATE BUTYRATE, TYPICAL EXAMPLES

	Superior Heat Resistance	Superior Shock Resistance
Specific Gravity	1.19 to 1.21	1.16 to 1.18
Impact strength ft. lbs. per in.	0.6 to 1.1	2.7 to 4.3
Tensile strength psi	5500 to 6900	2400 to 3500
Rockwell hardness	M65 to M71	M23 to M47
Heat distortion point °F.	179 to 208	125 to 144
Water absorption after 24-hr. im. %	1.6 to 2.0	1.3 to 1.4
Dielectric strength, 1/8", short time, v/m	250 to 400	250 to 400
Volume resistivity ohms-cm.	10^{10} to 10^{12}	10^{10} to 10^{12}
Coefficient of thermal expansion	11 to 17×10^{-5}	11 to 17×10^{-5}
Elongation %	51 to 63	71 to 79
Dielectric constant		
60 cycles	3.5 to 6.4	3.5 to 6.4
1 megacycle	3.2 to 6.2	3.2 to 6.2
3000 megacycles	2.95	2.95
Power factor		
60 cycles	0.01 to 0.04	0.01 to 0.04
1 megacycle	0.01 to 0.04	0.01 to 0.04
3000 megacycles	0.031	0.031

TABLE XVI

ETHYL CELLULOSE, THREE MAIN TYPES

	Injection Moulding Type	Extrusion Type	Low. Temp. Type
Density	1.08 to 1.18	1.08 to 1.18	1.08
Tensile strength psi	5000 to 12,000	2000 to 8000	7000 to 8000
Coefficient of thermal expansion	10 to 14×10^{-5}	10 to 14×10^{-5}	6×10^{-5}
Elongation %	1 to 15	15 to 100	6 to 8
Modulus of elasticity psi	—	—	$.9\times10^{5}$
Heat distortion point °F.	120 to 200	120 to 210	160 to 170
Water absorption after 24-hour immersion %	1.2 to 1.8	1.2 to 1.8	1.6
Dielectric strength (v/m on 020 sheet)	1400 to 1800	1400 to 1800	1500
Volume resistivity ohms-cm.	1 to 10×10^{12}	1 to 10×10^{12}	—
Dielectric constant			
1000 cycles	3.0 to 3.8	—	3.8
1 megacycle	3.2 to 3.7	—	3.7
Power factor			
1000 cycles	0.008 to 0.015	—	—
1 megacycle	0.001 to 0.02	—	—

chloride and vinyl acetate are good and typical examples. Theoretically, the vinyl resins are derivatives of vinyl alcohol, and, as we shall see later, a plastic material can be derived from vinyl alcohol itself which possesses somewhat unusual and remarkable properties.

A. *Vinyl Chloride*

Monomeric vinyl chloride is a limpid liquid of low boiling point (−14°C.), first discovered by Regnault as early as 1838.[21] Under suitable conditions of heat and pressure, and in the presence of catalysts, vinyl chloride is transformed to a white fluffy polymer. This fluffy polymer can be moulded under heat and pressure to give rigid sheets which are characterized by their horny nature and low moisture absorption and inertness to a variety of chemical solvents, as well as their resistance to flame. For many applications, however, this material is too hard for use and possesses the disadvantage of becoming brittle at relatively high temperatures. It was found, however, by the use of suitable plasticizers, that the hard horny product could be transformed to an elastic material, the exact properties of which depend on the amount and nature of the plasticizer added. A considerable literature now exists on the various compounds of polyvinyl chloride, marketed under the trade names of Vinylite and Geon (formerly Koroseal).

By the correct choice of compounding ingredients, materials of great flexibility at sub-zero temperatures but of satisfactory high temperature characteristics can be prepared. The electrical properties of such materials are largely dependent on the chemical nature of the plasticizer employed; and, since the best plasticizers are esters having high dipole moments, the electrical properties are very sensitive to temperature and frequency changes. Because of the usefulness of this type of material, more work on the properties under various conditions has been done than with most other plastics, and much useful data have been accumulated. Some average values for different types of materials are given for comparison purposes in Table XVII.

TABLE XVII

VINYL CHLORIDE—AVERAGE VALUES

Test	Flame Resistant Grade	High Insul. Resistance Grade	Cold Resistant Grade
Specific gravity	1.37	1.34	1.39
Tensile strength psi	2500	2300	2500
Modulus of elasticity psi	1100	1100	1400
Water absorption after 24-hour immersion %	1.35	0.40	0.30
Elongation %	300	350	325
Dielectric strength v/mil	350–800	400–850	325
Volume resistivity ohms-cm.	10^{11}–10^{13}	10^{13}–10^{15}	10^{8}–10^{11}
Dielectric constant			
60 cycles	6.4	6.2	
1000 cycles	4.2	4.9	
Power factor			
60 cycles	0.15	0.097	
1000 cycles	0.115	0.109	

The plasticized polyvinyl chlorides have found wide use as primary insulation for low frequency application such as drop wire, twisted telephone wire, and for cable jacketing purposes. Because of the high dielectric constant and high power factor, and the dependence of these properties on temperature and frequency, the plasticized polyvinyl chlorides are not usually employed in high frequency circuits.

B. *Copolymers of Vinyl Chloride and Vinyl Acetate.*

By the introduction of relatively small amounts of vinyl acetate into the polyvinyl chloride chain, a copolymer is obtained whose properties are distinctly different from those of polyvinyl chloride or polyvinyl acetate. It can be made into rigid plastic shapes or, by suitable plasticization, into elastomers.

The rigid varieties are noted for their strength, resistance to chemicals, non-inflammability, and dimensional stability. They are also highly resistant to moisture. Their chief drawback, however, is a relatively low softening point. (See Table XVIII.)

By the use of suitable compounding ingredients, as with the polyvinyl chlorides, tough elastomeric materials can be obtained. Compounds suitable for use as primary insulation or as the protective outer jacket of cables are available. Table XIX shows average values to be expected for typical materials.

TABLE XVIII

COPOLYMERS OF VINYL CHLORIDE AND VINYL ACETATE, AVERAGE PROPERTIES OF RIGID MATERIALS

Modulus of elasticity	35 to 41×10^{14}
Coefficient of thermal expansion	6.9×10^{-5}
Rockwell hardness	M60 to M80
Specific gravity	1.30 to 1.45
Heat distortion point °F.	140 to 150
Water absorption after 24-hour immersion %	0.05 to 0.15
Dielectric strength v/mil	400
Volume resistivity ohms-cm.	$>10^{14}$
Dielectric constant	
60 cycles	3.26
1000 cycles	3.21
1 megacycle	3.08
Power factor	
60 cycles	0.008
1000 cycles	0.031
1 megacycle	0.014

As in the vinyl chloride type materials, the electrical properties are very temperature and frequency dependent.

C. *Polyvinyl Acetate*

Liquid vinyl acetate is a mobile colorless, nontoxic liquid boiling at 73°C., obtained by the reaction of acetylene and acetic acid in the presence of a catalyst. When heated to 100°C. the liquid polymerizes to a colorless resinous material; reaction is accelerated by suitable catalysts such as oxygen or peroxides. Polyvinyl acetate was early suggested as a plastic base for lacquers and other types of coatings. Owing to the very low softening point of polyvinyl acetate and the development of newer synthetic plastics having superior electrical and physical properties, technical interest in polyvinyl acetate is rapidly diminishing; its use is limited to the adhesive leather finishing and lacquer field. It is, therefore, not necessary to consider this material further.

D. *Polyvinyl Alcohol*

By hydrolysis of polyvinyl acetate, the water soluble polyvinyl alcohol is obtained. Vinyl alcohol itself has never been isolated and the polymer can only be formed by this somewhat roundabout method. Polyvinyl alcohol is a remarkable material which finds a certain outlet in the plastics industry in the manufacture of hose and tubing for the conveyance of oils, greases, and other organic solvents, to which polyvinyl alcohol is remarkably resistant. Such materials are sold in the trade under the names of PVA and Resistoflex. Because of the hygroscopicity of the material and its chemical constitution, the electrical properties of the material are not such as to make it of interest for electrical insulation.

E. *Polyvinyl Acetals*

By suitable chemical treatment of polyvinyl acetate, a series of resins finding wide application has been prepared. The polyvinyl butyrals, when plasticized, form tough, impact resistant, adhesive interlayer materials whose principal use is in the manufacture of safety glass. The electrical properties of this material are not very good. The corresponding formal, however, shows quite interesting properties and is coming into use as an insulating material for electrical wires. Material of this type, because of its moisture re-

TABLE XIX

COPOLYMERS OF VINYL CHLORIDE AND VINYL ACETATE, AVERAGE VALUES OF ELASTOMERIC MATERIALS FOR PRIMARY INSULATION, ETC.

Test	Primary Insulation	Low Temp. Type	General Jacket Type
Tensile strength psi	3000	1700	2200
Specific gravity	1.32	1.22	1.24
Elongation %	200	250	250
Low temperature brittleness	−17°C.	−46°C.	−38°C.
Dielectric strength v/mil	300	300	300
D.C. resistivity	6.2×10^{6}	5.1×10^{4}	1.34×10^{6}
Dielectric constant			
60 cycles	6.0	7.4	7.6
1000 cycles	4.8	6.7	6.9
Power factor			
60 cycles	0.105	0.064	0.060
1000 cycles	0.120	0.067	0.098

Here again the electrical properties of the materials depend very largely on the temperature and frequency at which measurements are made.

sistance, solvent resistance, temperature stability, abrasion resistance, flexibility, toughness, and dielectric strength in very thin films, has led to its development and adoption as a superior insulating enamel on magnet wire (Formex—see Table XX).

The trade names for these plastics are Butacite, Butvar, Saflex, Saflex TS, Vinylite X, etc.

3. POLYVINYLIDENE CHLORIDE

One of the latest thermoplastic resins developed is polyvinylidene chloride. This plastic is non-inflammable, dimensionally stable, and has excellent mechanical properties. Electrically, however, it is not very good, and the material, therefore, finds its chief applications in the field of plastic tubing where resistance to a wide variety of chemicals is desired; and, because of its unusually high tensile strength in oriented films, as a textile fibre. See Table XXI.

4. ACRYLIC ACIDS

The history of acrylic resins, now so famous for their use as transparent astrodomes in aircraft use and as potential windshields for the post-war car, goes back to 1843 when Redtenbacher [22] isolated acrylic acid. In 1873 Caspary and Tollens [23] prepared the methyl ethyl and allyl ester and observed that a clear, transparent material was obtained from the allyl derivative on standing at room temperature. To Rohm and von Pechman, however, must go the chief credit for bringing the acrylate resins to successful commercial production and use.[24] Rohm secured his first patent for a rubber substitute by the use of polyacrylates in 1912.[25] The polymer most widely used is the methyl methacrylate known in the trade as Plexiglas and Lucite. It is marketed in the form of moulding powder available in a wide range of colors and in cast sheets, some of quite large size. Although finding its widest use as a transparent

TABLE XX

FORMEX

Tensile strength ft. lbs. per in.	9000 to 12,000
Coefficient of thermal expansion	7.7×10^{-5}
Modulus of elasticity	26 to 10^{5}
Rockwell hardness	M80 to M90
Elongation %	4 to 11
Specific gravity	1.2 to 1.3
Heat distortion point °F.	160 to 170
Water absorption after 24-hour immersion%	0.6 to 1.3
Dielectric strength v/m (In very thin films, may be as high as 1600 v/m)	300 to 600
Dielectric constant	
60 cycles	3.6 to 3.7
1000 cycles	3.3
1 megacycle	3.0
Power factor	
60 cycles	0.007
1000 cycles	0.01
1 megacycle	0.02

TABLE XXI

POLYVINYLIDENE CHLORIDE

Impact strength ft. lbs. per in.	2 to 8
Tensile strength unoriented psi	4000 to 7000
Coefficient of thermal expansion	15.8×10^{-5}
Elongation %	10 to 40
Rockwell hardness	M50 to M65
Heat distortion point °F.	150 to 180
Specific gravity	1.68 to 1.75
Dielectric strength, 1/8", short time, v/m	400
Volume resistivity ohms-cm.	10^{14} to 10^{16}
Dielectric constant	
60 cycles	3 to 5
1000 cycles	3 to 5
1 megacycle	3 to 5
Power factor	
60 cycles	0.03 to 0.08
1000 cycles	0.03 to 0.15
1 megacycle	0.03 to 0.05

material where good light transmission properties are required, the physical characteristics of methyl methacrylate polymers are such that they are potentially useful in other applications. Although not outstanding in electrical properties, considerable quantities of polymer are nevertheless used in electrical systems (See Table XXII).

From Table XXII, it will be seen that the material is unsuitable for electrical systems where high frequencies are being handled, and where power loss must be kept to a minimum, but it is greatly superior to the thermosetting type of plastic. Recently, however, the availability of polystyrene in clear rod and sheet form has tended to lessen the use of the methacrylate material. There are many other commercial acrylates and methacrylates available, but data available on their electrical properties are rather meager.

5. Nylon

The work of Carothers [26] on synthetic linear polyamides laid the foundation for the development of a new class of synthetic polymers, the so-called nylons. These materials owe their chief fame to their fibre-forming characteristics, but very recently even wider applications for these materials have been found by the development of Nylon FM-1, an injection moulding material. Filaments of nylon are prepared by spinning molten polyamide through special spinnerets and then cold-drawing the resultant thread. This process lines up the long-chain molecules so that they become oriented and, in this position, very high tensile strengths can be obtained. Like polyethylene, nylon has a sharp melting point and, just above the melting point, the material is very fluid, thus necessitating very careful design of injection moulding nozzles.

Although nylons so far produced are not very good electrically compared with hydrocarbon plastics, the high softening point of the material makes it attractive for use where the electrical properties are not critical. Table XXIII, giving physical and electrical properties of a general purpose nylon and a special FM-1 injection moulding material, will illustrate the point.

TABLE XXII

Methyl Methacrylate Polymers, Average Properties

Tensile strength psi	7000
Modulus of elasticity in tension psi	3 to 5×10^5
Rockwell hardness	M70 to M90
Specific gravity	1.18 to 1.20
Coefficient of thermal expansion °C.	7 to 9×10^{-5}
Elongation %	1
Heat distortion point °F.	125 to 165
Water absorption after 24-hour immersion %	0.2 to 0.3
Dielectric strength, 1/8", short time, v/m	500
Volume resistivity ohms-cm.	$>10^{15}$
Dielectric constant	
60 cycles	3.6
1000 cycles	3.0
1 megacycle	2.8
Power factor	
60 cycles	.05 to .07
1000 cycles	.05 to .07
1 megacycle	.02 to .03
Arc resistance	Good

6. Polyvinyl Carbazole

The development of a plastic material from vinyl carbazole originated in the laboratories of the I. G. Farben Industries. More recently this material has been made in this country. Although somewhat brittle, the material, because of its high resistance to deformation by heat and also its electrical properties, has found use as a substitute for mica in many radio frequency blocking and by-pass capacitors. It can also be used in high frequency and DC operation, such as filter, pulse, and coupling. Data on this material are very meager, and the most reliable results are those obtained on actual tests on capacitors built with this material. The DC breakdown on one grade of the material is of the order of 3000 volts per mil on a sheet 2 mils in thickness. The dielectric constant is approximately 5.5, while the power factor at one megacycle is .016. The power factor drops off as the temperature increases, as would be expected from a material of this type. The material can be operated as high as 125°C. without adverse effect. Polyvinyl carbazole is resistant to aliphatic hydrocarbons, alcohols, and ethers, but is soluble in aromatic hydrocarbons such as benzene and chlorinated hydrocarbons such as chlorobenzene. Its extended use is believed limited owing to the inherently brittle

TABLE XXIII

NYLON

Test	General Purpose	FM-1
Impact strength ft. lbs. per in.	2.18 to 4.25	9.4
Tensile strength psi	9000	10,500
Modulus of elasticity in tension psi	3.25×10^5	3.25×10^5
Specific gravity	1.14	1.14
Coefficient of linear expansion per °C.	10.3×10^{-5}	5.7×10^{-5}
Melting point °C.	263	
Elongation %	83	54
Dielectric strength v/m	380 to 400	400
Volume resistivity ohms-cm.	10^{13}	10^{13}
Dielectric constant		
60 cycles	3.8 to 7.0	3.8
1000 cycles	4.0 to 20	4.0
Power factor		
60 cycles	0.018 to 0.13	.018
1000 cycles	0.05 to 0.11	.05

nature of the plastic and its somewhat difficult processing.

7. HYDROCARBON PLASTICS

Probably the most important plastic materials from the standpoint of electrical properties are the hydrocarbon type materials. Because of their essentially non-polar nature in general, they possess low dielectric constants and good power factors over a wide range of temperatures and frequencies. We shall discuss here three of the most common of these materials, namely polyethylene, polystyrene, and polybutene.

A. *Polyethylene*

One of the most recently discovered of the thermoplastic materials is polyethylene. Ethylene itself is a colorless gas at room temperature; it has been known for several hundred years. Experiments involving the polymerization of ethylene have been described in scientific literature as far back as 1797; and, during the nineteenth century, thermal polymerization of ethylene in the presence of metallic chlorides was found to yield paraffinic oils together with other materials. It was not until about 1935, however, that the British found a method for producing a solid polymer from ethylene. By heating the gas at pressures above 1000 atmospheres and temperatures up to 400°C. with a carefully controlled amount of oxygen, Fawcett, Gibson, and Perrin [27] produced a solid, tough, waxy polymer having a somewhat sharp melting point around 105°C. In 1941, the material became available in the U. S. A. and, at the present moment, considerable quantities of this very valuable plastic are being produced.

Polyethylene is remarkable in that it has the lowest dielectric constant of any commercially available plastic, as well as a correspondingly low power factor, wide temperature range of usefulness, and low specific gravity. At the present time, polyethylene finds its principal outlet as the primary insulation of high and ultra-high frequency transmission lines. Among its adverse properties must be classed its flammability, its somewhat sharp melting point as distinguished from the usual gradual softening of other thermoplastic materials, its high coefficient of expansion, and its tendency to become contaminated when brought in contact with other materials having poorer electrical properties. Owing to its comparative newness and improvements that have been and are being made both in its physical and electrical properties, it is possible to include only an indication of its properties. Being a pure hydrocarbon analogous to paraffin wax, polyethylene suffers oxidation when heated in air. This

TABLE XXIV

POLYETHYLENE, AVERAGE PROPERTIES

Tensile strength psi	1700
Coefficient of thermal expansion	10.5×10^{-5}
Rockwell hardness	R13
Specific gravity	.92
Elongation %	500
Heat distortion point °F.	140
Water absorption after 24-hour immersion %	0.01
Dielectric strength v/m 3 to 15 mils thick	1000 to 1500
Volume resistivity ohms-cm.	10^{17}
Dielectric constant	
60 cycles	2.22 to 2.26
1000 cycles	2.22 to 2.26
1 megacycle	2.22 to 2.26
1000 megacycles	2.22 to 2.26
3000 megacycles	2.22 to 2.26
Power factor	
60 cycles	.0002
1000 cycles	.0002
1 megacycle	.0002
1000 megacycles	.0002 to .0004
3000 megacycles	.0003 to .0005

oxidation is accompanied by an increase in power factor, and it is customary, therefore, to add a small percentage of an anti-oxidant to the material to offset this tendency. Table XXIV lists the average properties of polyethylene.

B. *Polystyrene*

Polystyrene, discovered by Simon as early as 1839,[28] is the oldest synthetic organic plastic. Details of its formation from monomeric styrene are given in a paper previously published in this journal.[1] Recent developments in improved methods of manufacture and considerable experience gained through the preparation of large quantities of styrene necessitated by the synthetic rubber program, together with the inherently valuable properties of the material itself, have made this plastic one of the most potentially useful materials developed. It has low specific gravity, excellent dimensional stability, outstanding electrical properties, and remarkably low moisture absorption; it can be readily fabricated by conventional methods. Polystyrene is known in the trade under the names of Loalin, Bakelite, Styron, Lustron, Distrene, etc. The physical properties of a typical commercial material are given in Table XXV.

Because of the remarkable and outstanding electrical properties of polystyrene, more determinations of its electrical characteristics with frequency and temperature variations have been made than on any other thermoplastic material. It has been found that these values remain substantially constant over a wide range of conditions.

Dielectric strength 1/8", short time,	500 to 700 v/m
Arc resistance	120 to 135 seconds

TABLE XXV

POLYSTYRENE, PHYSICAL PROPERTIES OF A TYPICAL COMMERCIAL MATERIAL

Tensile strength psi	5500 to 7000
Coefficient of thermal expansion	6 to 8×10^{-5}
Modulus of elasticity in flexure psi	4.0 to 4.7×10^{5}
Specific gravity	1.05
Heat distortion point °F.	165 to 190
Water absorption after 24-hour immersion %	0.00
Elongation %	2 to 3
Rockwell hardness	M80 to M90

Because of the chemical nature of the plastic, whereby depolymerization occurs when the arc takes place, the material will withstand repeated arcing without substantial lowering of its arc resistance. The surface and volume resistivity of polystyrene are so high that it is difficult to obtain accurate measurements, but values for the surface resistivity are greater than 10^{16} ohms-cm., while its volume resistivity is 10^{17} to 10^{19} ohms-cm.

Dielectric Constant

The following values for the dielectric constant have been obtained at room temperature:

60 cycles	2.55
1000 cycles	2.55
50,000 cycles	2.53
1 megacycle	2.53
20 megacycles	2.52

Measured at 60 cycles, with a potential gradient of 50 volts per mil, the dielectric constant with increasing temperatures is shown below:

Temperature °C.	Dielectric Constant
30	2.55
60	2.57
90	2.60
Power Factor	
60 cycles	0.0002
1000 cycles	0.00015
40,000 cycles	0.00022
1 megacycle	0.00016
20 megacycles	0.00028
1000 megacycles	0.00028

The two limiting properties of polystyrene are its relatively low softening point—165 to 190°F.—and its rigidity at room temperature. Recent developments have led to improvement in the heat resistance of polystyrene, some of which are potentially very important. Perhaps the first attempt at increasing the heat resistance of polystyrene was the formation of the "cross linked" polystyrenes prepared by adding small amounts of paradivinyl benzene to the styrene before polymerization.[29] A cross linked polystyrene so formed is infusible and insoluble in solvents, and on heating becomes rubbery, retaining its shape to a remarkable degree. Because of the thermosetting nature of the reaction, only cast rods or

forms can be prepared of this material, and it is necessary, therefore, subsequently to machine the rod to the correct size. Other cross linking agents such as complex crotonates and allyl derivatives have been suggested in this connection. At the present moment, cross linked polystyrene rod is marketed by the General Electric Company under the code No. 1421.

A different approach to the problem was made by the Monsanto Chemical Company, which developed a heat resistant polystyrene known as Styramic by the incorporation of chlorinated diphenyl to the polystyrene itself. This had the additional advantage of conferring a certain measure of flame resistance to the material. The electrical properties of Styramic, while very satisfactory for a number of applications, are not quite so good as those of pure polystyrene, the dielectric constant in particular being somewhat higher and having a value between 2.8 and 3.0. An even more recent development has been the manufacture of the polydichlorostyrenes which possess much higher softening points and whose electrical properties are comparable to those of pure polystyrene itself. At the present moment, no details concerning this material can be divulged. Undoubtedly, we shall see more of these improved materials as time goes on.

In an attempt to secure greater flexibility, the Dow Chemical Company investigated many copolymers of styrene with other materials and have marketed a material called Styraloy, which exhibits remarkable flexibility even at low temperatures and which is reasonably resistant to heat. It has excellent abrasion resistance and good resistance to organic solvents and oils. Its electrical characteristics, while reasonably good at low frequencies, exhibit a region of anomalous dispersion above a megacycle when the power factor reaches quite high values. For this reason, Styraloy is unsuitable for ultra-high frequency work where the lowest of power factors is necessary. At the present moment, Styraloy finds its greatest outlet as the primary insulation of ignition wiring. The properties of this material are shown in Table XXVI.

C. *Polybutene*

When isobutylene, a gas at ordinary temperatures, is treated with a catalyst such as boron trifluoride at comparatively low temperatures around −80°C., it can be polymerized to form a very rubbery material known as polybutene, vistanex, or oppanol. The exact molecular weight and physical properties of the polymer depend on the temperature at which the polymerization is carried out, and the product ranges from a soft, sticky liquid to a very tough rubber. The polymer has great resistance to strong chemicals and ozone, and its electrical properties are excellent. Owing to its cold flow, however, the material is seldom used by itself but is usually compounded with other materials to improve their properties. It has found wide use as a means of improving the viscosity index of lubricating oils and as hot melt adhesives when mixed with paraffin wax. The low temperature flexibility characteristics of

TABLE XXVI

STYRALOY

Tensile strength psi	1000
Specific gravity	.957
Heat distortion point °F.	150
Water absorption after 24-hour immersion %	0.2
Elongation %	200 to 250
Dielectric strength 1/8", short time, v/m	700
Volume resistivity ohms-cm.	10^{20}
Dielectric constant	
1000 cycles	2.5 to 2.6
1 megacycle	2.6 to 2.7
50 megacycles	2.5
Power factor	
1000 cycles	0.0007 to 0.0012
1 megacycle	0.0005 to 0.0010
50 megacycles	0.0035 to 0.0040
100 megacycles	0.0040 to 0.0050

TABLE XXVII

POLYBUTENE

Tensile strength psi	500
Elongation %	550
Low temperature flexibility	Not brittle at −78°C.
Heat distortion point	Below room temperature
Power factor	
60 cycles	.0003 to .0005
1000 cycles	.0003 to .0005
1 megacycle	.0003 to .0005
10 megacycles	.0004 to .0006
100 megacycles	.0004 to .0006
Dielectric constant	
60 cycles	2.25 to 2.35
1000 cycles	2.25 to 2.35
1 megacycle	2.25 to 2.35
10 megacycles	2.25 to 2.35
100 megacycles	2.25 to 2.35

the material are outstanding. Because of the variable nature of properties obtainable, depending on the molecular weight, it is not possible to give a very accurate picture of physical properties. However, material having a molecular weight of approximately 100,000 can be expected to have the characteristics indicated in Table XXVII.

Conclusion

The wealth of data and the variety of materials available are such that the design engineer can now choose his material for the desired characteristics with some certainty that he will be able to get such characteristics and, moreover, with the distinct possibility that he will have a choice of materials. In the foregoing summary, which of necessity is incomplete and rather fragmentary, we have attempted to collect and collate such data as will facilitate an intelligent selection of materials. But it cannot be too strongly emphasized that, in the present state of the plastic art, a mere tabulation of physical and electrical properties will not necessarily enable anyone to choose the right material for any particular application. There still exists and will always exist the need for the exercise of judgment and the use of the accumulated knowledge of the particular engineer having the problem under advisement.

For the highest degree of electrical performance where dielectric properties are of paramount importance, the pure hydrocarbons such as polystyrene and polyethylene are the best materials available, and it is extremely difficult to envisage materials that will show marked advantages on this point over these two plastics. Undoubtedly other plastics will be developed combining their excellent electrical properties with somewhat superior mechanical performance, and when such materials become available, they will find very ready application. Where cheapness, availability, and good overall physical properties are required for low frequency applications, and where only moderate electrical performance is required, the recent developments in the urea formaldehydes are to be considered. For large sheets of various thicknesses for panel mounting strips, etc., and where high dielectric strength, reasonably low moisture absorption, and reasonable electrical properties are required, Grade XXXP laminated plastics will be a logical choice.

Our ideas on this most important subject of the application of plastics will undoubtedly undergo many changes, some of them fundamental in nature, under the impact of new developments which have received such a stimulus due to war conditions, but the final consideration as to the ultimate potentialities will always be the intelligent use to which our knowledge is applied. Tremendous strides have been made in the last ten years in the plastics field, and it can be stated unhesitatingly that plastics are here to stay—our task is to determine their limitations and possibilities and to plan accordingly.

References

1. "The Polymerization of Styrene and Some Concepts of the Electrical Properties of Plastics" by A. J. Warner, *Electrical Communication*, **21** (1943), pages 180–193.
2. A. Baeyer, *Ber.* **5**, 25, 280, 1094 (1872).
3. Impact strength, A.S.T.M. D256–41T.
4. Tensile strength, A.S.T.M. D638–42T.
5. Modulus of elasticity in tension, A.S.T.M. D638–42T.
6. Rockwell hardness. Proposed Federal Specification for Organic Plastics. General Spec. (Method of Physical Tests), July 7, 1942.
7. Specific gravity, A.S.T.M. D71–27.
8. Heat distortion point, A.S.T.M. D648–41T.
9. Water absorption, A.S.T.M. D570–42.
10. Dielectric strength, A.S.T.M. D149–40T.
11. Volume resistivity, A.S.T.M. D247–38.
12. Dielectric constant, A.S.T.M. D150–42T.
13. Power factor, A.S.T.M. D150–42T.
14. Arc resistance, A.S.T.M. D495–42.
15. Schutzenberger, *Compt. rend.* **61**, 485 (1865).
16. Francimont, *Compt. rend.* **89**, 711 (1879).
17. Suida, *Monatsh.* **26**, 413 (1905).
18. G.P. 322,586.
19. F.P. 462,274.
20. B.P. 12,854, U.S.P. 1,188,376.
21. R. Regnault, *Ann. Chim.* (2) **69**, 157 (1858).
22. Redtenbacher, *Ann.* **47**, 113–48 (1843).
23. Caspary and Tollens, *Ann.* **167**, 247–252 (1873).
24. Rohm and von Pechman, *Ber.* **34**, 427–9 (1901); *Ber.* **34**, 573–4 (1901), etc.
25. G.P. 262,707, U.S.P. 1,121,134.
26. Carothers, *J.A.C.S.* **54**, 1566–69 (1932).
27. U.S.P. 2,153,553.
28. Simon, E., *Ann.* **31**, 267 (1839).
29. Staudinger and Heuer, *Ber.* **67**, 1164 (1934).

Erratum, Vol. 21, No. 4, 1944

Standard Telephones and Cables, Limited, London—60th Anniversary

The sentence at the end of page 213 and the beginning of page 214 should read:

"At the outbreak of war in September, 1939, the number of employees was 78 per cent greater than in 1928 and the business had increased by 112 per cent in the same period."

INTERNATIONAL TELEPHONE AND TELEGRAPH CORPORATION

Associate Companies in the Western Hemisphere

UNITED STATES OF AMERICA

INTERNATIONAL STANDARD ELECTRIC CORPORATION: *Manufacturer and Supplier of Communication and Other Electrical Equipment Through Licensee Companies Throughout the World; Exporter of Communication and Other Electrical Equipment*.....New York, N. Y.

FEDERAL TELEPHONE AND RADIO CORPORATION: *Manufacturer of Communication and Other Electrical Equipment*.....Newark, N. J.

COMMERCIAL CABLE COMPANY: *Trans-Atlantic Telegraph Service*.....New York, N. Y.

COMMERCIAL PACIFIC CABLE COMPANY: *Trans-Pacific Telegraph Service*.....New York, N. Y.

MACKAY RADIO AND TELEGRAPH COMPANY: *International and Ship-Shore Radio Telegraph Services; Supplies, Operates and Maintains Marine Communication and Navigational Equipment*..... New York, N. Y.

ALL AMERICA CABLES AND RADIO, INC.
All America Cables and Radio, Inc. maintains 67 Company-owned telegraph offices in 23 countries and islands throughout Central and South America and the West Indies.....New York, N. Y.

THE CUBAN ALL AMERICA CABLES, INC.: *United States-Cuba Telegraph Service*.....New York, N. Y.

ARGENTINA

*COMPAÑÍA STANDARD ELECTRIC ARGENTINA: *Manufacturer of Communication and Other Electrical Equipment* .....Buenos Aires

COMPAÑÍA INTERNACIONAL DE RADIO (ARGENTINA): *Radio Telephone and Telegraph Services*. Buenos Aires

SOCIEDAD ANÓNIMA RADIO ARGENTINA: *Radio Telegraph Service* .....Buenos Aires

COMPAÑÍA TELEFÓNICA ARGENTINA: *Telephone Operating System*.....Buenos Aires

COMPAÑÍA TELEGRAFICO-TELEFÓNICA COMERCIAL: *Telephone Operating System*.....Buenos Aires

UNITED RIVER PLATE TELEPHONE COMPANY, LIMITED: *Telephone Operating System*.....Buenos Aires

BOLIVIA

COMPAÑÍA INTERNACIONAL DE RADIO BOLIVIANA: *Radio Telephone and Telegraph Services*.....La Paz

BRAZIL

*STANDARD ELECTRICA, S. A.: *Manufacturer of Communication and Other Electrical Equipment*
Rio de Janeiro

COMPANHIA RADIO INTERNACIONAL DO BRASIL: *Radio Telephone and Telegraph Services*.....Rio de Janeiro
Additional Stations: São Salvador (Bahia), Belem, Curityba, Fortaleza, Natal, Porto Alegre, Recife.

COMPANHIA TELEFONICA PARANAENSE, S. A.: *Telephone Operating System* .....Curityba

COMPANHIA TELEPHONICA RIO GRANDENSE: *Telephone Operating System* .....Porto Alegre

CHILE

*COMPAÑÍA STANDARD ELECTRIC, S.A.C.: *Manufacturer of Communication and Other Electrical Equipment*
Santiago

COMPAÑÍA DE TELÉFONOS DE CHILE: *Telephone Operating System* .....Santiago

COMPAÑÍA INTERNACIONAL DE RADIO, S. A. (CHILE): *Radio Telephone and Telegraph Services*.....Santiago

CUBA

CUBAN TELEPHONE COMPANY: *Telephone Operating System* .....Havana

RADIO CORPORATION OF CUBA: *Radio Telegraph Service*
Havana

MEXICO

MEXICAN TELEPHONE AND TELEGRAPH COMPANY: *Telephone Operating Syste* .....Mexico City

PERU

COMPAÑÍA PERUANA DE TELÉFONOS LIMITADA: *Telephone Operating System*.....Lima

PUERTO RICO

PORTO RICO TELEPHONE COMPANY: *Telephone Operating System*.....San Juan

RADIO CORPORATION OF PORTO RICO: *Radio Telephone Service and Radio Broadcasting*.....San Juan

---

Associate Companies in the British Empire

*STANDARD TELEPHONES AND CABLES, LIMITED: *Manufacturer of Communication and Other Electrical Equipment* .....London, England
Branch Offices: Birmingham, Leeds, England; Glasgow, Scotland; Cairo, Egypt; Calcutta, India; Pretoria, South Africa.

*CREED AND COMPANY, LIMITED: *Manufacturer of Teleprinters and Other Communication Equipment*
Croydon, England

INTERNATIONAL MARINE RADIO COMPANY, LIMITED: *Supplies, Operates and Maintains Marine Communication and Navigational Equipment*..Liverpool, England

*KOLSTER-BRANDES LIMITED: *Manufacturer of Radio Equipment* .....Sidcup, England

*STANDARD TELEPHONES AND CABLES PTY. LIMITED: *Manufacturer of Communication and Other Electrical Equipment* .....Sydney, Australia
Branch Offices: Melbourne, Australia; Wellington, N. Z.

*Licensee Manufacturing and Sales Company of the International Standard Electric Corporation, New York, N. Y.