VOLUME XXIX JULY, 1950 NO. 3

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

50¢ per copy

$1.50 per Year

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the

American Telephone and Telegraph Company

195 Broadway, New York 7, N. Y.

Leroy A. Wilson, *President* — Carroll O. Bickelhaupt, *Secretary* — Donald R. Belcher, *Treasurer*

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are 50 cents each. The foreign postage is 35 cents per year or 9 cents per copy.

PRINTED IN U. S. A.

The Bell System Technical Journal

Vol. XXIX *July, 1950* *No. 3*

Principles and Applications of Waveguide Transmission

By GEORGE C. SOUTHWORTH

Under the above title, D. Van Nostrand Company, Inc. will shortly publish the book from which the following article is excerpted. Dr. Southworth is one of the leading authorities on waveguides and was one of the first to foresee the great usefulness that this form of transmission might offer. The editors of the Bell System Technical Journal are grateful for permission to publish here parts of the preface and the historical introduction and chapter 6 in its entirety.

Preface

Though it has been scarcely fifteen years since the waveguide was proposed as a practicable medium of transmission, rather important applications have already been made. The first, which was initiated several years ago, was in connection with radar. A more recent and possibly more important application has been in television where waveguide methods provide a very special kind of radio for relaying program material cross-country from one tower top to another. Already Boston and New York have been connected by this means and shortly Chicago and intervening cities will be added. Other networks extending as far west as the Pacific may be expected. It is reasonable to expect that these two applications will be but the beginning of a more general use.

Interest in the subject of waveguide transmission is not limited to commercial application alone. A comparable interest, perhaps less readily evaluated but nevertheless extremely important, lies in its usefulness in teaching important physical principles. For example there are many concepts that follow from the electromagnetic theory that, in their native mathematical form, may appear rather abstract. However, when translated to phenomena actually observed in waveguides, they become very real indeed. As a result, these new techniques have already assumed a place of considerable importance in the teaching of electrical engineering and applied physics both in lecture demonstrations and in laboratory exercises. It is to be expected that they will be used even more extensively as their possibilities become better appreciated.

Interest in waveguides has been greatly enhanced by the fact that they brought with them a series of extremely interesting methods of measurement, comparable both in accuracy and scope, with similar measurements previously made only at the lower frequencies. This extension of the range over which electrical measurements may be made has contributed also to neighboring fields of research. One early application led to the discovery of centimeter waves in the sun's spectrum. Another led to important new information about the earth's atmosphere. Still another contributed to the study of absorption bands in gases, particularly bands in the millimeter region. Also of great importance was its contribution to our knowledge of the properties of materials for it led at a fairly early date to measurements at higher frequencies than heretofore of the primary constants, permeability, dielectric constant and conductivity—all for a wide array of substances ranging from the best insulators to the best conductors and including many of the so-called semi-conductors. It is because this new art has already attained considerable stature and is already showing promise as an educational medium that this book has been prepared.

CHAPTER I

INTRODUCTION

1.5 Early History of Waveguides

That it might be possible to transmit electromagnetic waves through hollow metal pipes must have occurred to physicists almost as soon as the nature of electromagnetic waves became fully appreciated. That this might actually be accomplished in practice was probably in considerable doubt, for certain conclusions of the mathematical theory of electricity seemed to indicate that it would not be possible to support inside a hollow conductor the lines of electric force of which waves were assumed to consist. Evidence of this doubt appears in Vol. I (p. 399) of Heaviside's "Electromagnetic Theory" (1893) where, in discussing the case of the coaxial conductor, the statement is made that "it does not seem possible to do without the inner conductor, for when it is taken away we have nothing left on which tubes of displacement can terminate internally, and along which they can run."

Perhaps the first analysis suggesting the possibility of waves in hollow pipes appeared in 1893 in the book "Recent Researches in Electricity and Magnetism" by J. J. Thomson. This book, which was written as a sequel to Maxwell's "Treatise on Electricity and Magnetism," examined mathematically the hypothetical question of what might result if an electric charge should be released on the interior wall of a closed metal cylinder. This problem is even now of considerable interest in connection with resonance in hollow metal chambers. The following year Joseph Larmor examined as a special case of electrical vibrations in condensing systems the particular waves that might be generated by spark-gap oscillators located in hollow metal cylinders. A more complete analysis relating particularly to propagation through dielectrically-filled pipes both of circular and rectangular cross section was published in 1897 by Lord Rayleigh. Later (1905) Kalähne examined mathematically the possibility of oscillations in "ring-shaped" metal tubes. Still later (1910) Hondros and Debye examined mathematically the more complicated problem of propagation through dielectric wires. Transmission through hollow metal pipes was also considered by Dr. L. Silberstein in 1915.

As regards experimental verification, it is of interest that Sir Oliver Lodge as early as 1894 approached but probably did not quite realize actual waveguide transmission. In a demonstration lecture on electric waves given before the Royal Society, he used, as a source of waves, a spark oscillator mounted inside a "hat-shaped" cylinder. An illustration published later suggests that the length of the cylinder was only slightly greater than its diameter. There is no very definite evidence that the short cylinder functioned as a waveguide or that such a function was discussed in the lecture. Perhaps of greater significance were some experiments reported a year later by Viktor von Lang who used pipes of appreciable length and repeated for electric waves the interference experiment that had been performed for acoustic waves by Quincke some years earlier. Other similar experiments were later performed by Drude and by Weber.

About 1913 Professor Zahn of the University of Kiel became interested in this problem and assigned certain of its aspects to two young candidates for the doctorate, Schriever and Reuter by name. They had barely started when World War I broke out, and both left for the front. Zahn continued this work until he was called a year later. It is reported that by this time he had succeeded in propagating waves through cylinders of dielectric, but it is understood that he did little or no quantitative work. Reuter was killed at Champagne in the autumn of 1915, but Schriever survived and returned to complete his thesis in 1920, using for his source the newly available Barkhausen oscillator.

The contributions of Thomson, Rayleigh, Hondros and Debye, and Silberstein were, of course, purely mathematical. Those of von Lang, Weber, Zahn and Schriever were experimental, but they were of rather limited scope. The concept of the hollow pipe as a useful transmission element, for example as a radiator or as a resonant circuit, apparently did not exist at these early dates. Nothing was yet known quantitatively about attenuation, and little or nothing of the present-day experimental technique had yet appeared. At this time, the position of this new art was perhaps comparable with that of radio prior to the time of Marconi.

The history of waveguides changed abruptly about 1933 when it was shown that they could be put to practical use. Several patent applications were filed,[1] and numerous scientific papers were published. More recently a great many papers have appeared, too many in fact for detailed consideration at this time. Three of the earlier papers are mentioned in the footnote below.[2] Others will be referred to in the text that follows.

The writer's interest in guided waves stems from some experiments done in 1920 when such waves were encountered as a troublesome spurious effect while working with Lecher wires in a trough of water. In one case there were found, superimposed on the waves that might normally travel along two parallel conductors, other waves having a velocity that somehow depended on the dimensions of the trough. These may now be identified as being the so-called dominant type. In another case, the depth of water was apparently at or near "cut-off," and conditions were such that water waves in the trough gave rise to depths that were momentarily above cut-off, followed a moment later by depths that were below cut-off. This led not only to variations in power at the receiving end of the trough but also to variations in the plate current of the oscillator supplying the wavepower. Indeed these effects could be noted even when the wires were removed from the trough. These waves were recognized as being roughly like those described the same year by Schriever.[3]

Several years later this work was resumed and since that time a continued effort has been made to develop from fundamental principles of waveguide transmission a useful technique for dealing with microwaves. The earliest of these experiments consisted of transmitting electromagnetic waves through tall cylinders of water. Because of the high dielectric constant of water, waves which were a meter long in air were only eleven centimeters long in water. Thus it became possible to set up in the relatively small space of one of these cylinders many of the wave configurations predicted by theory. In addition it was possible, by producing standing waves, to measure their apparent wavelength and thereby calculate their phase velocity. Also by investigating the surface of the water by means of a probe,

[1] Reference is made particularly to U.S. Patents 2,129,711 (filed 3/16/33, 2,129,712 (filed 12/9/33), 2,206,923 (filed 9/12/34) and 2,106,768 (filed 9/25/34).

[2] Carson, Mead and Schelkunoff, "Hyper-frequency Waveguides—Mathematical Theory," *B.S.T.J.*, Vol. 15, pp 310–333, April 1936. G. C. Southworth, "Hyper-frequency Wave Guides—General Considerations and Experimental Results," *B.S.T.J.*, Vol. 15, pp 284–309, April 1936. Also "Some Fundamental Experiments with Waveguides," *Proc. I.R.E.*, Vol. 25, pp 807–822, July 1937. W. L. Barrow, "Transmission of Electromagnetic Waves in Hollow Tubes of Metal," *Proc. I.R.E.*, Vol. 24, pp 1298–1398, October 1936.

[3] The waves actually observed are now known as TE_{10} waves in a rectangular guide, while those described by Schriever are now recognized as TM_{01} waves in a circular guide.

the directions and also the relative intensities of lines of electric force in the wave front could be mapped. It is probable that certain of these modes were observed and identified for the first time.

Shortly afterwards, sources giving wavelengths in air of fifteen centimeters became available and the experimental work was transferred to air-filled copper pipes only 5 inches in diameter. At this time, a 5-inch hollow-pipe transmission line 875 feet in length was built through which both telegraph and telephone signals were transmitted. Measurements showed that the attenuation was relatively small. This early work, which was done prior to January 1, 1934, was described along with other more advanced work in demonstration-lectures and also in papers published in 1936 and 1937.[4]

It was recognized at an early date that a short waveguide line might, with suitable modification, function as a radiator and also as a reactive element. These properties were likewise investigated experimentally, and numerous useful applications were proposed. Descriptions may be found in the numerous patents that followed. These properties were also the subject of several experimental lectures given before the Institute of Radio Engineers and other similar societies by the writer and his associates during the years 1937 to 1939.[5] Included were demonstrations of the waveguide as a transmission line, the electromagnetic horn as a radiator, and the waveguide cavity as a resonator. An adaptation of the waveguide cavity was used to terminate a waveguide line in its characteristic impedance.

From the first, progress was very substantial and by the autumn of 1941 there were known, both from calculation and experiment, the more important facts about the waveguide. In particular, the reactive nature of discontinuities became the subject of considerable study, and impedance matching devices (transformers), microwave filters, and balancers soon followed. Also a wide variety of antennas was devised. Similarly, amplifiers and oscillators as well as the receiving methods followed.

As might be expected, a great many people have contributed in one way or another to the success of this venture. Particular mention should be made of the very important parts played by the author's colleagues, Messrs. A. E. Bowen and A. P. King, who, during its early and less promising period, contributed much toward transforming rather abstract ideas into practical equipment, much of which found important military uses immediately upon the advent of war. Also of importance were the parts played by the author's colleagues, Dr. S. A. Schelkunoff, J. R. Carson, and Mrs. S. P. Meade, who, in the early days of this work, provided a substantial segment of mathematical theory that previously was missing. During the succeeding years, Dr. Schelkunoff, in particular, made invaluable contributions in the form

[4] A description of one of the earlier lectures appears in the Bell Laboratories Record for March 1940. (Vol. XVIII, No. 7, p. 194.)

of analyses which in some cases indicated the direction toward which experiment should proceed and, in others, merely confirmed experiment, while, in still others, gave answers not readily obtainable by experiment alone. In the chapters that follow, the author has drawn freely on Dr. Schelkunoff, particularly as regards methods of analysis.

Beginning sometime prior to 1936, Dr. W. L. Barrow, then of the Massachusetts Institute of Technology, also became interested in this subject and together with numerous associates made very substantial contributions. No less than eight scientific papers were published covering special features of hollow-pipe transmission lines and electromagnetic horns. For several years the work being done at the Massachusetts Institute of Technology and at the Bell Telephone Laboratories probably represented the major portion, if not indeed the only work of this kind in progress, but with the advent of World War II, hundreds or perhaps thousands of others entered the field. For the most part, the latter were workers on various military projects. Starting with the considerable accumulation of unpublished technique that was made freely available to them at the outset of the war, they, along with others in similar positions elsewhere in this country and in Europe, have helped to bring this technique to its present very satisfactory state of development.

CHAPTER VI

A DESCRIPTIVE ACCOUNT OF ELECTRICAL TRANSMISSION

6.0 General Considerations

The preceding four chapters presented the more important steps in the development of the theory of electrical transmission, particularly as it applies to simple networks, wire lines, and waves in free space and in guides. For the most part, the analysis followed conventional methods and made use of the concise and accurate short-hand notation of mathematics. It had for its principal objective the derivation of a series of equations useful in the practical application of waveguides.

Closely associated with the theory of electricity and almost a necessary consequence of it are the numerous concepts and mental pictures by means of which we may explain rather simply the various phenomena observed in electrical practice Though extremely important, this aspect of the theory was not stressed before. Instead it was deferred to the present chapter where it could be considered by itself and from the purely qualitative point of view. It is hoped that this arrangement of material will be of special use to those who find it necessary to substitute for mathematical analysis, simple models to explain the phenomena which they observe in practice. It is believed that, for these people, this chapter together with a few key formulas taken from the earlier sections will be helpful in gaining a fairly satisfactory understanding of the practical aspects of waveguide transmission.

At the lower frequencies, the current aspect of electricity meets most of the needs and in comparison it is only occasionally that there is a need to discuss lines of electric and magnetic force. In waveguide practice, on the other hand, currents are usually not available for measurement and, although we recognize their reality, they necessarily assume a secondary role. In contrast with currents, we consider the fields present in a waveguide as very real entities and we attach a very great importance to their orientations as well as to their intensities.

6.1 The Nature of Fields of Force

As a suitable introduction to the discussion that follows, we shall review some of the fundamental properties of lines of electric and magnetic force and show pictorially the part that they play in transmission along an ordinary two-wire line.

The Electrostatic Field

As is well known, the concept of the electric field was devised by Faraday to explain the force action between charged bodies. According to his view there exist in the space between the charged bodies, lines or tubes of electric force terminating respectively on positive and negative charges attached to the bodies. These tubes of force are endowed with a tendency to become as short as possible and at the same time to repel, laterally, neighboring lines of force. Their direction at any point is purely arbitrary, but, by subsequent convention, the positive direction is taken from the positively charged body to the negative. This is such that a small positive charge (proton) placed in the field tends to be displaced in the positive sense while an electron tends to move in a negative direction. The force exerted on the unit charge is a measure of the magnitude of the electric intensity **E**. It is measured in volts per meter and, since it has direction as well as magnitude, it is a vector quantity.[1] Figure 6.1-1 illustrates in a general way the arrangement of lines of electrostatic force that are assumed to exist between two oppositely charged spheres. Also shown is a representative vector **E**.

The Magnetostatic Field

In the same way that Faraday provided a satisfactory explanation for the forces between charged bodies, so was he able to explain the forces be-

[1] Black-face type will be used when it seems desirable to emphasize the vector properties of quantities having direction as well as magnitude.

tween magnetized bodies. In the latter case, the two kinds of electrostatic charge are replaced by north-seeking and south-seeking magnetic poles respectively. Similarly the tubes of electric force are replaced by tubes of magnetic force. Roughly speaking, the two kinds of tubes are endowed with analogous properties. Because these magnetic lines are at rest, it is appropriate to speak of them as magnetostatic lines of force and consider them as being comparable but of course not identical with electrostatic lines already discussed. The force exerted on a unit magnetic pole is a measure of magnetic intensity **H**. Like its electric counterpart, it is a vector quantity. In the par-

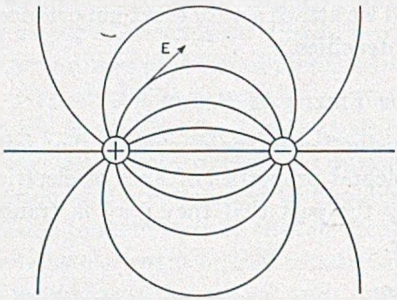


Fig. 6.1-1. Arrangement of lines of electrostatic force in the region between two oppositely charged spheres.

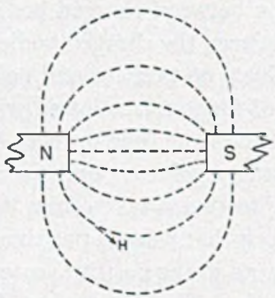


Fig. 6.1-2. Arrangement of lines of magnetostatic force in the region between two oppositely magnetized poles.

ticular system of units used in this text, it is measured in amperes per meter. Figure 6.1-2 illustrates the arrangement of the lines of magnetic force that are assumed to exist between two opposite magnetic poles.

Interrelationship of Electric and Magnetic Fields

As a result of the electromagnetic theory, there are certain properties with which we may endow lines of electric and magnetic force and thereby explain numerous phenomena of electrical transmission. This establishes a relationship between electric and magnetic fields that makes them appear at times as if they were different aspects of the same thing. They are as follows:

1. *Lines of magnetic force, when displaced laterally, induce in the space immediately adjacent, lines of electric force. The direction of the induced electric force is perpendicular to the direction of motion and also perpendicular to the direction of the original magnetic force. The intensity* **E** *of the induced electric*

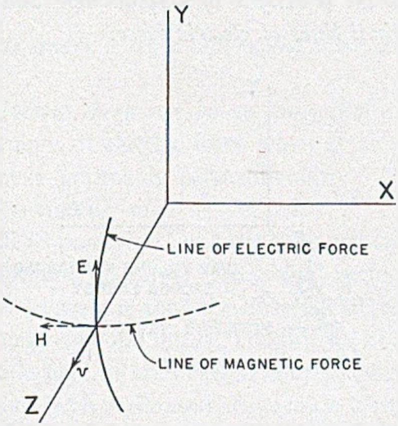


Fig. 6.1-3. Directions of electric vector E and magnetic vector H relative to the velocity v of motion of such lines.

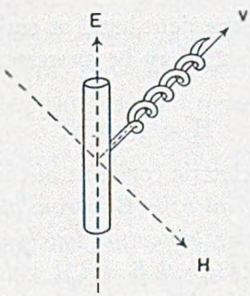


Fig. 6.1-4. Simple corkscrew rule for remembering the directions of E, H and v.

force is proportional to the velocity **v** *of displacement and proportional to the intensity* **H** *of the original lines of magnetic force.*

The directions of the vectors **v**, **E** and **H** are shown in Fig. 6.1-3. They are so related that, when **E** moves clockwise into **H**, it is as though a right-hand screw had progressed in the direction of **v** as shown in Fig. 6.1-4. A convenient short-hand notation used rather generally by mathematicians makes it possible to express these facts by the following vector equation:

$$\mathbf{E} = -\mu(\mathbf{v} \times \mathbf{H}) \tag{6.1-1}$$

The quantity μ is the magnetic permeability of the medium under consideration.

2. *Lines of electric force, when displaced laterally, induce in the immediately adjacent space lines of magnetic force. The direction of the induced magnetic force is perpendicular to the direction of motion and also perpendicular to the direction of the original electric force. The intensity* **H** *of the induced magnetic force is proportional to the velocity* **v** *of displacement and proportional to the intensity* **E** *of the original lines of electric force.*

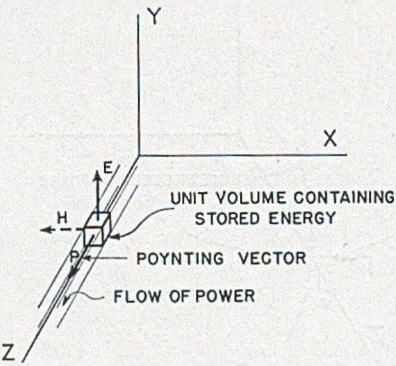


Fig. 6.1-5. Directions of the vectors E and H relative to the Poynting vector P in an advancing wave front.

Again Fig. 6.1-3 and also the right-hand or cork-screw rule apply. In the short-hand notation these facts may be expressed by the following vector equation:

$$\mathbf{H} = \epsilon(\mathbf{v} \times \mathbf{E}) \tag{6.1-2}$$

In this equation, ϵ is the dielectric constant of the medium.[2]

3. *When an electric field of intensity* **E** *is translated laterally, it together with its associated magnetic field* **H** *represents a flow of energy. The direction of the flow of energy is perpendicular to both* **E** *and* **H** *and is therefore in the direction of the velocity* **v**. *The magnitude of the energy flow per unit volume across a unit area measured perpendicular to* **v** *is proportional to the product of the electric intensity* **E** *and the magnetic intensity* **H**. *It may be designated by the vector* **P**.

The relative directions of the vectors **P**, **E**, and **H** are shown in Fig. 6.1-5. The energy per unit volume moves with a velocity expressed by

$$v = \frac{1}{\sqrt{\mu\epsilon}} \tag{6.1-3}$$

[2] The values of permeability μ and dielectric constant ϵ appearing in these equations are not the values found in most tables of the properties of materials. As here given μ is smaller than the usual value μ_r by a factor of 1.257×10^{-6} while ϵ is smaller than ϵ_r by a factor of 8.854×10^{-12}. The use of these special values leads to certain mathematical simplifications.

It therefore corresponds to a flow of power. In the notation just referred to, it may be expressed by the vector equation

$$\mathbf{P} = \mathbf{E} \times \mathbf{H} \tag{6.1-4}$$

4. *Lines of force exhibit the properties of inertia. They therefore resist acceleration.*

Other principles not quite so fundamental but nevertheless useful in application are:

5. *Lines of force are under tension and at the same time are under lateral pressure.*

6. *For perfect conductors there can be no tangential component of electric force.* That is to say, lines of electric force when attaching themselves to a perfect conductor must approach perpendicularly. This is substantially true also for common metals such as copper.

In passing it is well to point out that the first principle is really that by which the ordinary dynamo operates. The second is, for practical purposes, Oersted's Principle, if we assume that the lines of electric force are attached to charges flowing in near-by conductors. The third is known as the Poynting Principle. It has a wide field of application contributing very materially to the physical pictures of both radio and waveguide transmission. When applied to the very simple case of low frequencies propagated along a transmission line, it gives a result that is in keeping with the usual view that the power transmitted is equal to the product of the total voltage times the total current. The fourth principle is useful in explaining qualitatively how radiation from an antenna takes place. The usefulness of these four principles will be made more evident by the examples that follow.

6.2 Transmission of Power along a Wire Line

Direct Current

According to the Poynting concept, one may think of an ordinary dry cell as two conductors combined with chemical means for producing a continuous supply of lines of electric force. This need not be counter to the accepted views concerning electrolysis, for we may think of these lines of force as being attached to ionic charges incidental to dissociation. As long as the cell is on open circuit, these lines of electric force remain in a static condition in which many are grouped in the neighborhood of the terminals of the cell as shown in Fig. 6.2-1(a). In this state of equilibrium, the forces of lateral pressure are balanced by the forces of tension. There is no motion and hence no flow of power. For an ordinary dry cell such as used in flashlights, the electric intensity **E** will depend on the spacing of electrodes, but it may be as much as 200 volts per meter. If we attach to the dry cell two parallel wires spaced perhaps a centimeter apart with their remote ends open, electrostatic lines will be communicated to the wires, thereby providing a distribution roughly like that shown in Fig. 6.2-1(b). Except at the moment of contact, there is no motion of the lines of electric force and therefore no magnetic field and, accordingly, there can be no flow of power. The final configuration is to be regarded as the resultant of the forces of tension and lateral pressure. The electric intensity, **E**, measured in volts per meter at any point along the line, may be altered at will, merely by changing the spacing.

If, next, we close the remote end of the line by substituting a conducting wire for the particular line of force shown as a heavy line in Fig. 6.2-1(c), the adjacent lines of electric force will collapse on the terminating conductor,

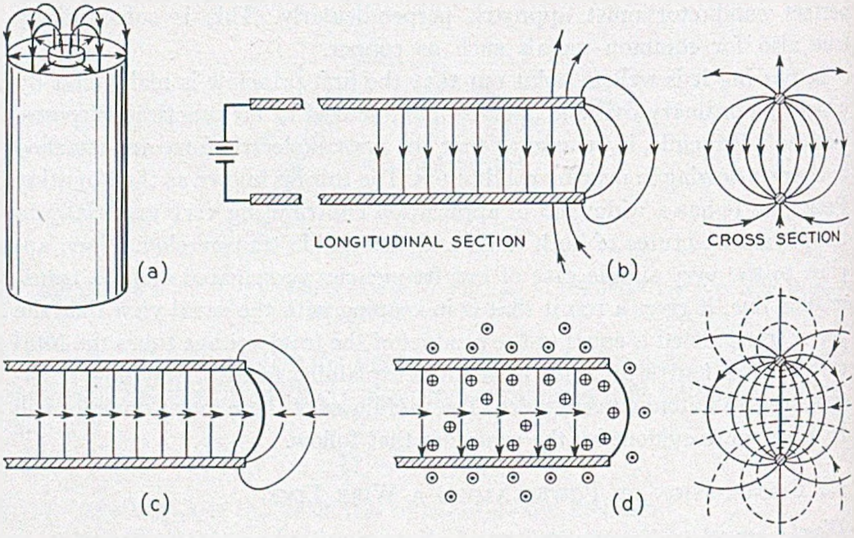


Fig. 6.2-1. Lines-of-force concept applied to the transmission of d-c power along a wire line.

as opposing charges unite. This removes the lateral pressure on the neighboring lines with the result that the whole assemblage starts moving forward. Each line of force meets in its turn the fate of its forerunners, thereby delivering up its energy to the resistance as heat. As soon as the lateral pressure at the cell is relieved, chemical equilibrium is momentarily destroyed and more lines of force are manufactured to fill the gaps of those that have gone before. All of this is, of course, at the expense of chemical action.

According to the electromagnetic theory, as set forth in the second principle, this is but a part of the story of transmission. We must add that the motion of the lines of electric force from the dry cell toward the resistance gives rise in the surrounding space to lines of magnetic force in accordance with Equation 6.1–2 and furthermore the two fields together give rise to component Poynting vectors representing power flow. Each component vector has a magnitude at any point equal to the product of the electric and magnetic intensities there prevailing and a direction at right angles to the two component forces in accordance with Equation 6.1-3. This is illustrated in Fig. 6.2-1(d).

Since the fields reside largely outside the conductors, we conclude that the principal component of power flow is through the space between the wires and not through the wires themselves. If, in the case cited above, there is appreciable resistance in the connecting wires, then we may expect that there will be a small component of energy flowing into the wires to be dissipated as heat. To account for this, we may picture lines of electric force

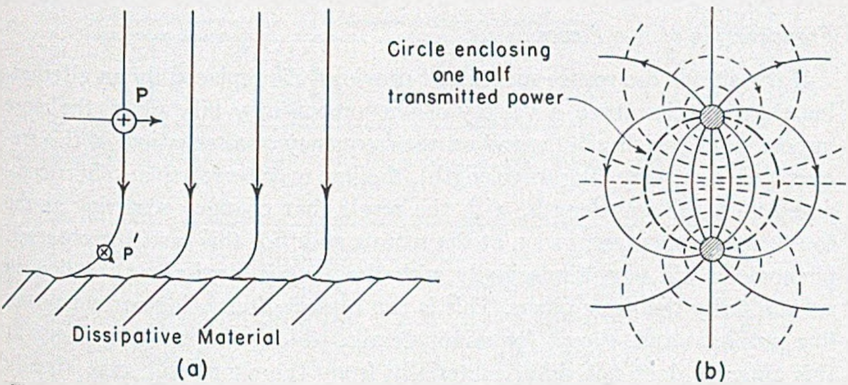


Fig. 6.2-2. Fields of electric and magnetic force and also direction of power flow in the vicinity of conductors. (a) Magnified view showing power flow along a single dissipative wire. (b) Cross-sectional view of parallel-wire line.

which in the immediate vicinity of the conducting wire lag somewhat behind the portions more remote. This is illustrated by Fig. 6.2-2(a) which shows a highly idealized and greatly enlarged section of the field in the immediate vicinity of one of the two dissipative conductors. The very small component of power flowing into the conductor is designated as the vector P' to distinguish it from the much greater power P which we shall assume is being propagated parallel to the conductor.[3]

The magnetic field associated with two cylindrical conductors consists of circles with centers on the line joining the two conductors, whereas the electric field consists of another series of circles orthogonally related to the

[3] For all metals from which conducting lines are ordinarily made, the component of power flowing into the conductor is extremely small compared with the power flowing parallel to its surface. In Fig. 6.2-2(a) therefore, we should regard vector P' as greatly exaggerated in magnitude relative to that of vector P.

first, and having centers on a line at right angles to the first as shown in Fig. 6.2-2(b). The total flow of power through any plane set up perpendicular to the wires is found by adding up the various component products of E and H from the boundaries of the wires to infinity. The method by which this is carried out is outside of the scope of this chapter, but, as already pointed out, it leads to the same result as obtained by multiplying together the total voltage and the total current. There are two results of this integration that are of special interest. (1) In the case of two parallel cylinders, one-half of the total power flows through the space enclosed by a circle drawn about the wire spacing as a diameter [see Fig. 6.2-2(b)]. The remaining half extends from this circle on out to infinity. (2) Since both the electric and magnetic intensities are greatest in the neighborhood of the wire, most of the total power flow takes place in the immediate vicinity of the wire.

Transmission of A-c Power

If the simple d-c source mentioned previously is replaced by an alternating electromotive force, a variety of phenomena may take place, the more important of which will depend on the frequency of alternation. If this frequency is low (very long wavelength), the line may be relatively short compared with the wavelength, with the result that changes occurring at the source may appear very soon at the remote end. For this case, the observed phenomena will vary sinusoidally with time everywhere along the line, in substantially the same phase. This is the typical alternating-current power line problem[4] and, except for minor details, which we shall not discuss at this time, it does not differ materially from the simple d-c case already covered.

If, on the other hand, the frequency is high (short wavelength), the line may be regarded as being *electrically long*, with the result that sinusoidal changes occurring at the source may not have traveled very far before the direction of flow at the source has changed. The over-all result in extreme cases may become very complicated indeed; for, wavepower may not only be reflected from the remote end of the line but, if there are sharp bends in the line or abrupt changes in spacing, it may be reflected from these points also. The phenomenon observed is usually referred to as *wave interference* and it often leads to *standing waves*. Though described above as complicated, there are many cases where the results of wave interference may be sufficiently simple to be readily visualized. Practical difficulties of various kinds may arise from these effects, but they may also serve very useful purposes. In fact, a substantial portion of our microwave technique is based on wave

[4] The wavelength corresponding to a frequency of 60 cycles per second is five million meters. A commercial power line having a length as great as 100 miles is therefore but 0.03 wavelength long. It is said to be *electrically short*.

interference. Certain specific examples will be discussed later, but first we shall discuss a somewhat simpler case.

The Infinite Line

Let us take, for discussion, a uniform two-wire line that is infinitely long. Waves launched on such a line are assumed to be propagated to infinity. There are no reflected components and hence no wave interference. If the frequency is very high, the forerunners of the lines of force sent out by the source will not have traveled very far when the emf at the source will have reversed its direction. This gives rise at the source to a second group of lines

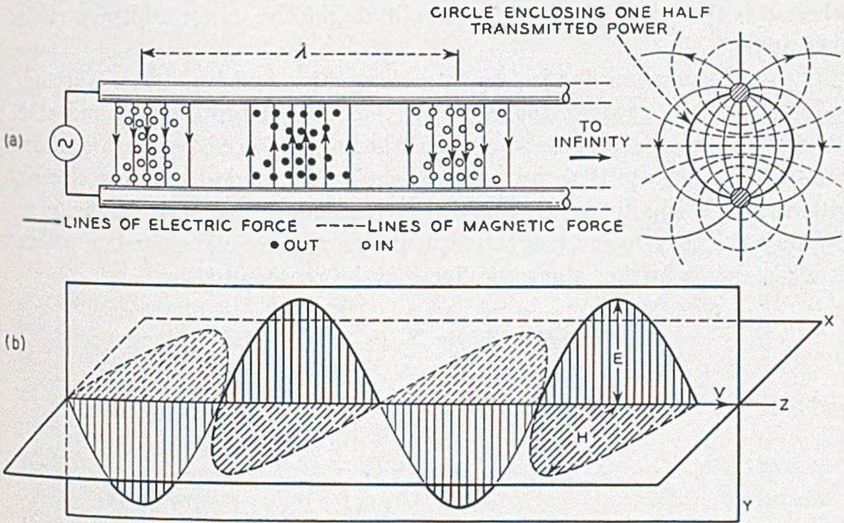


Fig. 6.2-3. (a) Arrangement of lines of electric and magnetic force in both the longitudinal and transverse sections of an infinitely long transmission line. (b) Space relationship between electric vector E and magnetic vector H as observed in a plane containing the two conductors.

of force exactly like the first except oppositely directed. This, in turn, will be followed by a third group identical with the first and a fourth identical with the second and so forth until equilibrium is reached. Because the lines of electric force are in motion, we must expect them to be accompanied by lines of magnetic force. Both are of equal importance. Therefore it is not correct to refer to either alone as a distinguishing feature of the wave. Both components are shown in cross section at the right in Fig. 6.2-3(a).

The distance between successive points of the same electrical phase in a wave is known as the wavelength λ. It depends on the frequency f of alternation and the velocity of propagation v; $\lambda = v/f$. The velocity of propagation in turn depends on the nature of the medium between the two wires. For air, the velocity v_a is substantially 300,000,000 meters per second (186,000 mi per sec). For other media $v = v_a/\sqrt{\mu_r \epsilon_r}$. Thus it will be seen that, by replacing the air normally found between the two wires of a transmission line by another medium such as oil ($\epsilon_r = 2$ and $\mu_r = 1$), the wavelength will be reduced by a factor of $1/\sqrt{2}$.

If A_0 is the maximum amplitude reached by the oscillating source during any cycle, the amplitude at any time t, measured from an arbitrary beginning, may be expressed by the equation

$$A = A_0 \sin (\omega t + \phi) = A_0 \sin \left(\frac{2\pi}{\lambda} vt + \phi \right) \tag{6.2-1}$$

where ϕ is the initial phase of the amplitude relative to an arbitrary reference angle.

If the transmission line is free from dissipation and we choose a datum point in a plane at right angles to the direction of propagation and at a distance far enough from the source that the lines of force have had an opportunity to conform to the wire arrangement and if we designate the electric intensity at this point as E_0 and the corresponding magnetic intensity as H_0, then the electric and magnetic intensities at other corresponding points at a distance z further along the line may be represented by

$$E = E_0 \sin \frac{2\pi}{\lambda} (z - vt)$$

and

$$H = H_0 \sin \frac{2\pi}{\lambda} (z - vt) \tag{6.2-2}$$

These equations are the trigonometric representations of an unattenuated sinusoidal wave of electric intensity and magnetic intensity traveling in a positive direction along the z axis. They are plotted in the yz and xz planes of Fig. 6.2-3(b). An electromagnetic configuration similar to the above but traveling in the opposite direction is given by

$$E = E_0 \sin \frac{2\pi}{\lambda} (z + vt)$$

and

$$H = H_0 \sin \frac{2\pi}{\lambda} (z + vt) \tag{6.2-3}$$

These equations may be further confirmed by plotting arbitrary values on rectangular-coordinate paper. In an infinite line the magnetic intensity **H** and the electric intensity **E** are in the same phase as shown in Fig. 6.2-3.

If the wave is subject to an attenuation of α units per unit distance, possibly due to resistance in the wires, the corresponding components of E and H are equally attenuated. Either component may be expressed by an equation of the type

$$E = E_0 e^{-\alpha z} \sin \frac{2\pi}{\lambda} (z - vt) \tag{6.2-4}$$

This is a very special form of certain equations appearing in Sections 3.2 and 3.3.

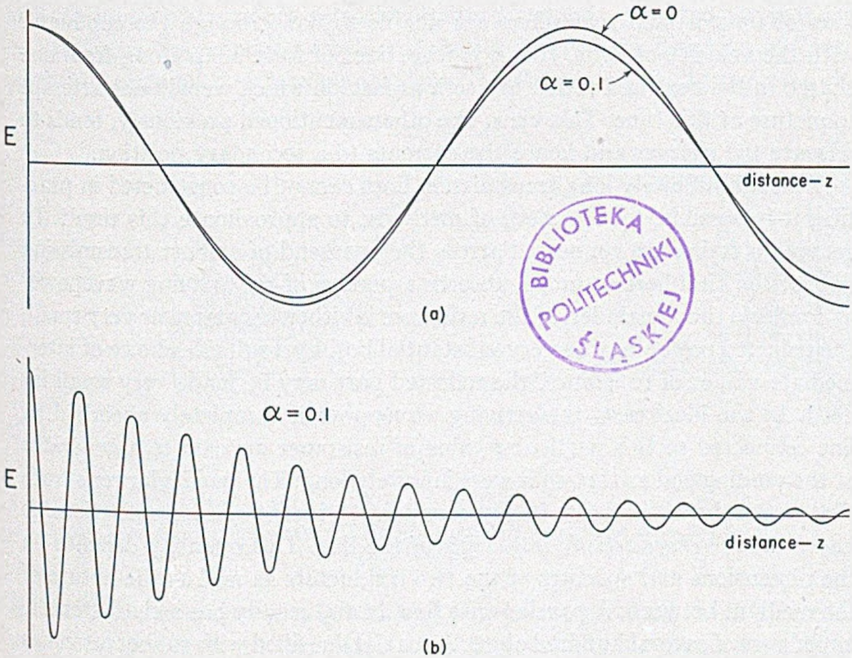


Fig. 6.2-4. Effect of attenuation on an advancing wave front.

If the attenuation is negligible, then $\alpha = 0$ and the term $e^{-\alpha z}$ will be unity. Equation 6.2-4 will then reduce to 6.2-2. If, on the other hand, the attenuation is considerable, the product of α times z will increase rapidly with distance, and the factor $e^{-\alpha z}$ will have the effect of reducing the electric intensity E prevailing at various points along the line. Figure 6.2-4(a) illustrates the variation, with distance, of the electric intensity E for an unattenuated wave $\alpha = 0$. There is included for comparison purposes the case, $\alpha = 0.1$. Figure 6.2-4(b) shows the effect of this rate of attenuation on waves that have traveled for some distance. It is significant that moderate amounts of attenuation have little or no effect on wavelength.

At low frequencies, conductor loss is often the principal cause of attenuation. At high frequency, this loss may be still more important[5] and in addition there may be losses in the medium around the two conductors. The latter is particularly true when the conductors are supported on insulators or are embedded in insulating material. There may also be losses due to lines of force that detach themselves from the wires and float off into the surrounding space (radiation). All three lead to attenuation and may be expressed in terms of an equivalent resistance. They are amenable to calculation for certain special cases.

According to one view of electricity, the individual charges to which lines of force attach themselves are unable to flow through the conductor with the velocity of light. If this is true, lines of force snap along from one charge to the next in a rather mysterious fashion which we will not attempt to picture at this time. This view, like others mentioned previously, tends to relegate the charges and hence the currents to a secondary position.

Although infinitely long transmission lines cannot be constructed in practice, it is possible, by a variety of methods, to approximate this result. In general, a resistance connected across the open end of a short transmission line, of the kind here assumed, absorbs a portion of the arriving wavepower and reflects the remainder. If the resistance is either very large or very small, the reflected power may be very substantial but, by a suitable choice of intermediate values of resistance, the reflected part may be made very small indeed. In the ideal case, the arriving wavepower is completely absorbed. A line connected to this particular value of resistance appears to a generator at the sending end as though it were infinitely long. The particular resistance that can replace an infinite line at any point, without causing reflections, is known as the *characteristic impedance* of the line. This quantity depends on the dimensions and spacings of the two conductors as well as the nature of the medium between. A parallel-wire line, in air, usually has a characteristic impedance of several hundred ohms. A coaxial line filled with rubber often has a characteristic impedance of a few tens of ohms. A line having characteristic impedance connected at its receiving end is said to be *match-terminated*.

Reflections on Transmission Lines

If the transmission line ends in a termination other than characteristic impedance, or if there are discontinuities, due to impedances connected either in series or in shunt with the line, reflections of various kinds will occur.[6] Much of the practical side of microwaves has to do with these reflections.

[5] The losses in most conductors increase with the square root of the frequency.

[6] At the higher frequencies, reflections may also occur at points where the wire spacing changes abruptly. In some instances abrupt changes in wire diameter may be sufficient to cause reflection. These discontinuities may be regarded as changes in characteristic impedance.

A particularly simple form of reflection occurs when the high-frequency transmission line is terminated in a transverse sheet of metal of good conductivity, as for example, copper. An arrangement of this kind is shown in Fig. 6.2-5. As it is difficult to represent a wave front moving toward the reflecting plate, we shall substitute an imaginary thin slice or section of the electromagnetic configuration. A slice of this kind is shown in Fig. 6.2-5(a).

Experiment shows that, at the boundary of the nearly perfect reflector, the transverse electric force E is extremely small. This is consistent with the sixth principle set forth in the previous section which states that there can be no tangential component of electric force at the boundary of a perfect conductor. The result actually observed can be accounted for if it is assumed that the reflecting conductor merely reverses the direction of lines of electric force as they become incident, thereby giving rise to two sets of

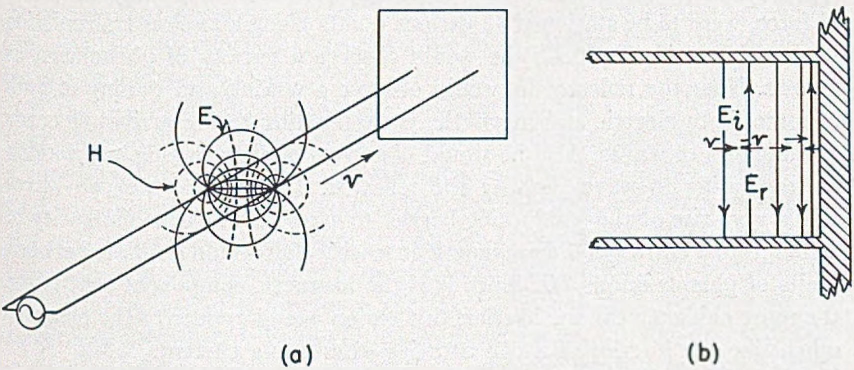


Fig. 6.2-5. (a) Propagation of an electromagnetic wave along a two-wire line terminated by a large conducting plate. (b) Representative lines of force reflected by the conducting plate.

lines of force as shown in Fig. 6.2-5(b), one of intensity $E_i = E$ directed downward in the figure and moving laterally toward the metal sheet (incident wave) and the other of intensity $E_r = -E$ directed upward and moving away from the metal sheet (reflected wave). Accordingly the resultant electric intensity at the surface is zero.

If the reflector is non-magnetic, the magnetic intensity **H** will be unaffected by the reflecting material. We find by applying the right-hand rule of Fig. 6.1-4 that the electric intensity $\mathbf{E}_r = -\mathbf{E}$ when combined with **H** constitutes a wave that must travel in a negative direction of **v**. This wave may be represented by Equation 6.2-3. In a similar way the Poynting vector which before reflection is represented by $\mathbf{P} = \mathbf{E} \times \mathbf{H}$ now takes the form $\mathbf{P} = (-\mathbf{E} \times \mathbf{H})$. The negative sign according to the right-hand rule of Fig. 6.1-4 shows that the power approaching the conductor is reflected back upon itself. If **E** and **H** are respectively equal in magnitude before and after incidence, the reflection is perfect, and the coefficient of reflection is said to be unity. Bearing in mind that $\mathbf{H}_i = \epsilon(\mathbf{v} \times \mathbf{E})$ before reflection and $\mathbf{H}_r = \epsilon(-\mathbf{v} \times -\mathbf{E})$ after reflection, it is evident that the direction of the magnetic intensity has been unchanged by the process of reflection and that the resultant magnitude at the surface of the metal is $|H_i| + |H_r| = 2|H|$. Thus we see that, at the moment of reflection from a metallic surface, the resultant electric force vanishes and the resultant magnetic force is doubled.

The reflection of waves at the end of the line naturally gives rise to two oppositely directed wave trains. This is a well-known condition for standing waves. Though a complete discussion of standing waves calls for the mathematical steps taken in Section 3.6, there are certain qualitative results that may be deduced from relatively simple reasoning. Some of these deductions will be made in the paragraphs that follow.

If an observer, endowed with a special kind of vision for individual lines of force, were to be stationed at various points along a lossless transmission line as shown in Fig. 6.2-5, he would observe a variety of phenomena as follows. Near the reflector he would observe a waxing and waning of lines of force, both electric and magnetic, corresponding to the arrival of crests and hollows of waves. Also he would observe a similar waxing and waning corresponding to waves leaving the reflector. The sum of the two waves would give rise at the conducting barrier to a resultant electric intensity of zero and to a corresponding magnetic intensity that would oscillate between limits of plus or minus $2H$. Since it is the magnetic component that is the the more evident near the barrier, this region would appear to the observer much like the interior of a coil carrying alternating current.

If the observer were to pass along the line to a point one-eighth wavelength to the left of the reflector, the distance up to the reflector and back would then be a quarter wave and he would then find that at the moment that a wave crest (maximum intensity) was passing on its way toward the reflector a point on the wave corresponding to zero intensity would be returning from the reflector. Adding the corresponding electric and magnetic intensities at this point, he would observe that the electric intensity would not always be zero but instead it would oscillate between limits of plus or minus $\sqrt{2}\,E$. Similarly the corresponding magnetic intensity would no longer oscillate between limits of plus or minus $2H$, but instead it would never reach limits greater than plus or minus $\sqrt{2}\,H$. Thus at this point the electric and magnetic components would have the same average intensity.

If the observer were to move farther along the line, stopping this time at a distance of one-fourth wavelength to the left of the metal plate, the total electrical distance to the barrier and back again would be a half wavelength and he would now find that at the time a crest passed on its way toward the reflector a hollow (maximum negative intensity) would be passing on its return journey. This time, the resultant electric intensity would oscillate between limits of plus or minus $2E$, and the resultant magnetic intensity would be zero at all times. To this observer then, this quarter-wave point on the line would have many of the characteristics of the interior of a condenser charged by an alternating voltage.

If our observer were to move another one-eighth wave farther along the line, he would note that the resultant electric and magnetic forces would again be equal. Proceeding on to a point one-half wavelength from the metal reflector, he would observe that, at the time crests (maximum positive intensity) were passing on their way toward the reflector, hollows would be returning, and accordingly upon examining the resultant electric intensity he would find it to be zero at all times, whereas the corresponding magnetic intensity would be oscillating between limits of plus or minus $2H$. At this point along the line, he would be unable to distinguish his electrical environment from that prevailing at the metal boundary. The half-wave line, therefore, has had the effect of translating the metal barrier to another point in space a half wave removed.

If the observer were to continue still farther along the line, he would pass, alternately, points where the resultant electric force is zero and other points where the resultant magnetic force is zero. It is important to note that at points in a standing wave where the magnetic force is a maximum, the electric force is a minimum and at points where the electric force is a maximum, the corresponding magnetic force is a minimum. It is customary to call the points of minimum E (or H) "mins," though the term *node* is sometimes substituted. Points of maximum E (or H) are known as "maxs" with the term *loop* as its alternative. If the observer were to measure current and voltage along the line, he would find that points of maximum voltage correspond to maximum E and that points of maximum current correspond to maximum H.

An examination of the energy associated with the incident and reflected waves shows that, except for minor losses not to be considered here, there is as much energy led away from the reflector as is led up to the reflector, and that there is associated with the standing wave a stored or resident energy. The regular arrangement of nodes and loops along a standing wave with minima at half-wave intervals is a very important characteristic, for such points may be located very accurately experimentally, and accordingly wavelength may be measured with considerable precision.

If, instead of terminating the wire line in a large conducting plane assumed previously, it is terminated in a relatively thin cross bar as shown in Fig. 6.2-6, the reflection will assume a somewhat more complicated form. First of all, the thin cross bar will intercept, initially at least, only a portion of the total wave front. The particular lines of force arriving along a plane containing the two wires will be the first to be reflected and they will behave at reflection much like those already discussed, whereas those outside the plane of the two wires will not be intercepted initially by the thin cross bar but instead will advance for a short distance beyond the end of the line before their forces of tension bring them to rest. These outlying lines of force are represented by the lines designated as *c* in Fig. 6.2-6. After the first lines of force have been reflected, lateral pressure will be removed from those adjacent, with the result that they will close in and collapse on the conductor at a slightly later time than their neighbors. One over-all result of this process is to make the effective length of such a line slightly greater than the true length. Effects of this kind are observed in practice and they are referred to as *fringing*. Discrepancies between the wavelength as measured in the last section of line where fringing may take place and that measured between other minima along the same line are usually small but

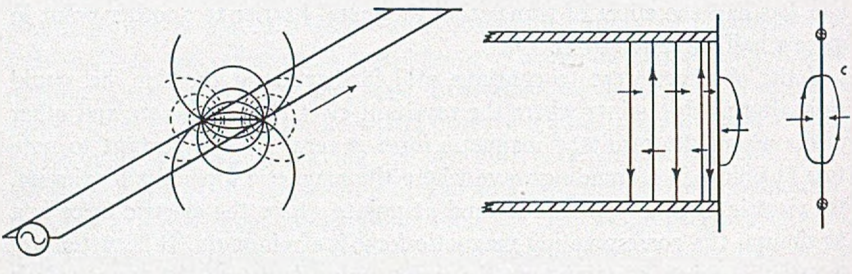


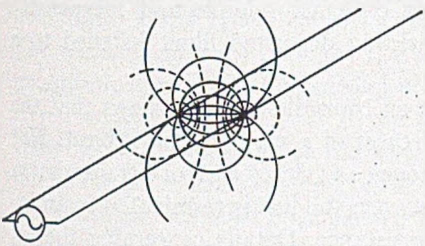
Fig. 6.2-6. (a) Representative transmission line terminated by a conductor of finite dimensions. (b) Nature of reflection by a finite conductor.

they are nevertheless measurable. It is also true that, as the wave front approaches a limited barrier of this kind, some of its energy continues on into the space beyond and is lost as radiation. In general, the smaller the barrier, the larger will be the losses.

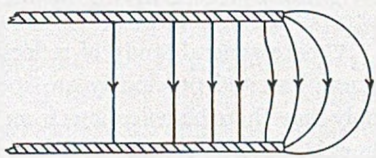
Consider next a line open at its remote end, as shown in Fig. 6.2-7. In this case, none of the lines of force of the advancing wave is intercepted by a conductor, with the result that a very considerable number momentarily congregate near the end of the line and, because of inertia, they extend into the space beyond as suggested by Fig. 6.2-7(b). This process continues until forces of tension in the lines, still clinging fast to the ends of the wires, bring the assemblage temporarily to rest. At this moment, there is no magnetic component; for v, in the relation $\mathbf{H} = \epsilon(\mathbf{v} \times \mathbf{E})$, is zero while the corresponding electric intensity is approximately $2E$. The lines of electric force, being momentarily at rest, represent energy stored in the electric form.

This static situation is extremely temporary, for the tension momentarily created in the lines of electric force soon forces the configuration as a whole to move backward. As the wave front gets under way, the magnetic force H increases in magnitude in accordance with the relation $\mathbf{H} = \epsilon(\mathbf{v} \times \mathbf{E})$.

The fact that the wave front extends momentarily for a short distance beyond the physical end of the line and requires time to come to rest and get into motion in the reverse direction implies inertia or momentum in the wave front. This is the inertia referred to in the fourth principle mentioned in Section 6.1. In this form of reflection, fringing is usually very evident, and because of fringing we may have an apparent reflection point that is considerably beyond the end of the wires. Thus the distance from the end of the wires back to the first voltage minimum is much less than the quarter wave that otherwise might be expected.



(a)



(b)

Fig. 6.2-7. (a) Transmission along a line open at the remote end. (b) Nature of reflection from open end.

It is generally true that processes of reflection in which fringing takes place are usually attended by considerable amounts of radiation. This suggests that in the process of reflection some of this extended wavepower detaches itself from the parent circuit and is lost. Experience shows that this lost power may be greatly enhanced by separating the two wires or by flaring their open ends. The so-called half-wave dipole, so familiar in ordinary radio, is but a transmission line in which the last quarter-wave length of each wire has been flared to an angle of 90 degrees. If we wish to minimize radiation, we follow a reverse procedure and reduce the spacing between the two parallel wires. This also reduces fringing, for we find that the measured distance from the ends of the wires to the first voltage minimum is now more nearly a quarter wave.

It is of interest to compare reflections taking place at the open end of a transmission line with those at a closed end. When a wave front becomes incident upon a perfect conductor, the electric force vanishes. At the same time, the lines of magnetic force, though effectively brought to rest, are momentarily doubled in intensity. The energy is predominantly magnetic, and the type of reflection may be regarded as inductive. When the wave is reflected from the ideal open-end line, a reverse situation prevails. The lines of magnetic force momentarily vanish while lines of electric force, though brought to rest, are doubled in intensity. At this moment the energy is predominantly electrostatic, and the reflection may be considered as being capacitive.

When a line is terminated in a sheet of metal of good conductivity such as copper or silver, reflection is almost perfect. If the sheet is a poor conductor such as lead or German silver, most of the incident power will still be reflected; but if a semi-conductor, such as carbon, is used as a reflector, a perceptible amount of the incident power will be absorbed. It is interesting also that the penetration into all metals at the time of reflection is very slight, for relatively thin sheets seem to serve almost as well as thick plates. It is therefore possible to use as reflectors extremely simple and inexpensive materials, for example, foils or electrically deposited films fastened to a cheaper material such as wood.[7]

A more general study of reflections on transmission lines shows that the examples cited previously are special cases of a very general subject. Not only may there be reflections from the open and closed ends of a transmission line, but there may be reflections also when the line is terminated in an inductance, in a capacitance, or in a resistance. Details concerning the reflections that may be observed from various combinations of these three impedances are discussed in connection with Fig. 3.6-3. The outstanding results of these discussions may be summarized for the ideal case as follows:

1. A pure inductance (positive reactance) connected at the end of a transmission line always leads to a reflection coefficient having a magnitude of unity. The standing wave resulting from this reflection will be characterized by the following: (a) If the terminating inductance is infinitely large (reactance of positive infinity), the reflection will be identical with that from an ideal open-end line, and the distance to the nearest voltage minimum will be a quarter wave. [See Fig. 3.6-3(a).] (b) If the inductance is finite but very large, the distance to the nearest voltage minimum, as measured toward the generator, will be somewhat greater than a quarter wave. [See Fig. 3.6-3(b).] (c) If the inductance is reduced progressively toward zero (reactance zero), the distance to the same voltage minimum will approach one-half wavelength. In this limiting case, another voltage minimum will appear at the end of the line. [See Fig. 3.6-3(c) and 3.6-3(d).]

2. A pure capacitance (negative reactance) connected at the end of a

[7] One convenient and inexpensive form of reflector is a kind of building paper coated with copper or aluminum foil. Moderately good reflectors can also be made by covering wood with a special paint containing finely divided silver in suspension (Du Pont's 4817). Most aluminum paints are unsatisfactory for this purpose.

transmission line also leads to a reflection coefficient having a magnitude of unity. In this case, the resulting standing wave will be characterized as follows: (a) If the capacitance is zero, (reactance equal to minus infinity), the reflection will correspond to that from the open end of a transmission line, and a voltage minimum will be found at a distance of a quarter wave from the end. [See Fig. 3.6–3(g).] (b) If the capacitance is increased from zero to a small finite value, the distance to the nearest voltage minimum will be somewhat less than a quarter wave. [See Fig. 3.6-3(f).] (c) If the capacitance is increased progressively toward infinity (reactance zero), the distance to the nearest voltage minimum will approach zero. [See Figs. 3.6-3(e) and 3.6-3(d).] The limiting condition, in which the terminating capacitance is zero, is comparable with that in which the termination is an infinitely large inductance.

3. If a pure resistance is connected at the end of a transmission line, the magnitude of the reflection coefficient varies with the resistance chosen. The relations are such that: (a) If the terminating resistance is infinite, the magnitude of the reflection coefficient will be unity and its sign will be positive. [See Fig. 3.6-3(h).] (b) If the terminating resistance approaches the characteristic impedance of the line, the distance to the nearest voltage minimum will remain constant, but the magnitude of the reflection coefficient will approach zero. [See Figs. 3.6-3(i) and 3.6-3(j).] (c) If the terminating resistance is made less than characteristic impedance, the sign of the reflection coefficient will be reversed, and, as the terminating resistance approaches zero, its magnitude will approach unity. [See Figs. 3.6-3(k) and 3.6-3(l).]

When the terminating resistance is infinite, the reflection is comparable with that in an ideal open-end line, and the nearest voltage minimum will be found at a distance of a quarter wave. When the terminating resistance is zero, the reflection is comparable with that in a closed-end line, and the voltage minimum will appear at the end of the line and also at a point one-half wave closer to the generator. If the line is terminated in a pure resistance of intermediate value, the voltage minima of such standing waves as may be present will be found at the end of the line for all values of the resistance that are less than characteristic impedance and a quarter wave removed from the end of the line for all values greater than characteristic impedance. When the terminating resistance equals characteristic impedance, there is no standing wave.

If, instead of terminating the line considered above in an inductance coil or in a capacitance or a resistance, we assume that it continues indefinitely into a mass of material having either a conductivity or a dielectric constant different from that of air, similar reflections may take place at the surface. A particular example is shown in Fig. 6.2-8. In general, a part of the wave-power arriving at the surface will be reflected and a part will be transmitted.

One may picture a portion of the Faraday tubes of force turned back at the interface while the remainder continue into the second medium. If one were to reverse the direction of transmission and consider wavepower transmitted from the second medium back into the first, a similar partial reflection would be noted. In both cases the part turned back and returned to the source may be regarded as a reactive component since no energy is really lost. In a similar way, the transmitted component, since it is not returned to the source, may be regarded as a resistive or dissipative component.

If the medium into which wavepower is transmitted is a perfect insulator, the transmitted wave will continue indefinitely except as attenuated by the

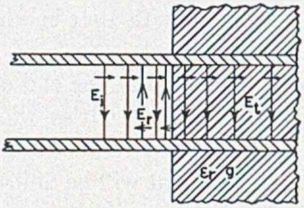


Fig. 6.2-8. Reflection and transmission of lines of force incidental to a change of medium along a transmission line.

wires along which it is guided. Its wavelength, λ, in the dielectric will be less than the wavelength, λ_0, in air as expressed by the relation

$$\lambda = \frac{\lambda_0}{\sqrt{\epsilon_r}}$$

If the second medium is somewhat conducting, the wave will be further attenuated, the rate of attenuation being related in a rather complicated way not only to the conductivity of the second medium b t to its dielectric constant and permeability as well. Thus far in microwave pract.ce, little practical use has been made of materials having permeabilities very different from unity. However, considerable use has been made of materials having various dielectric constants, ϵ_r, and conductivities, g. Sometimes these take the form of plates placed across a waveguide transmission line. Examples will appear in Section 9.8.

If a thin sheet of insulating material having a dielectric constant, ϵ_r, and conductivity of zero is placed across a two-wire transmission line, the percentage of power reflected is given approximately by

$$q_w = \frac{\pi l}{\lambda_0} (\epsilon_r - 1) \tag{6.2-5}$$

A thin sheet of this kind is approximated when wires carrying very high frequencies pass through the glass walls of a vacuum tube. If the glass thickness, l, is small compared with the wavelength in air, λ_0, the power reflected by the glass envelope will likewise be small.

Sometimes it is not feasible to reduce the wall thickness sufficiently to avoid serious reflections. In these instances it may be possible to make the thickness one-half wavelength as measured in glass whereupon the wave reflected from one face of the plate will be approximately equal in amplitude to that from the other face and, since they are separated by one-half wavelength, they tend to cancel.

Another case of practical interest is that in which the line is terminated in a plate of very special dielectric constant ϵ_r, conductivity g_1, and thickness l. This is followed by a second plate of nearly infinite conductivity. This arrangement is shown in longitudinal section in Fig. 6.2-9. By a proper choice of constants, the combination may be made a good absorber of wave-

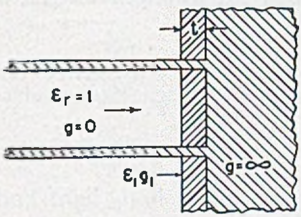


Fig. 6.2-9. A transmission line terminated in a conductor coated with a special materia such that all of the incident wave power is absorbed.

power. It will therefore be substantially reflectionless It may be shown that to satisfy this requirement

$$\lambda_0 = \frac{\sqrt{\epsilon_r}}{15\pi g_1} \tag{6.2-6}$$

and

$$l = \frac{1}{60\pi g_1(2n - 1)} = \frac{\lambda_0}{4\sqrt{\epsilon_r}\,(2n - 1)} \tag{6.2-7}$$

where n is any integer. One common example is that in which $n = 0$. The plate is then a quarter wave thick as measured in the medium.[8] A reflectionless plate of this kind when placed at the end of a transmission line appears to the source as though the line were terminated in its characteristic impedance. Devices incorporating this principle are sometimes used as match terminators for waveguides.[9]

[8] A more complete discussion of this problem was published in 1938 by G. W. O. Howe, "Reflection and Absorption of Electromagnetic Waves by Dielectric Strata," *Wireless Engr.*, Vol. 15, pp 593–595, November 1938.

[9] Plates of this kind may be made very simply by mixing carbon with plaster in varying proportions until the right combination is reached.

When a two-wire transmission line assumes the coaxial form, the lines of electric force are radial and lines of magnetic force are coaxial circles. The directions of these two components obey the right-hand rule. (See Fig. 6.2-10.) Since the wave configuration is completely enclosed except for a small exposure at each end, radiation from this type of line can be made very small.

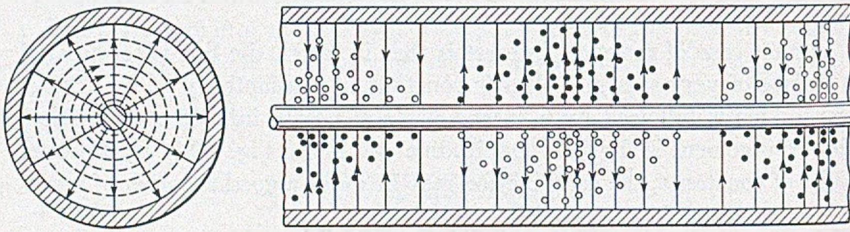


Fig. 6.2-10. Arrangement of lines of electric and magnetic force associated with transmission along a coaxial arrangement of conductors.

6.3 Radiation

Electromagnetic waves, including both light and radio waves, are not unlike the waves that are guided along wire lines. Their difference is largely a matter of environment. In one case they are attached to wires while in the other they have presumably detached themselves from some configuration of conductors and are spreading indefinitely into surrounding space. We shall present in this section one of several possible pictures of the launching of radio waves from a transmission line. Like other verbal pictures drawn in this chapter, it should be regarded as highly qualitative.

Assume a two-wire line with one end flared as shown in Fig. 6.3-1. If at some point to the left there is a source of wavepower, there will flow from left to right along the line a sinusoidal distribution of lines of electric and magnetic force not unlike that shown in Fig. 6.2-7. In order to simplify our illustration, we shall single out for examination two representative lines of electric force *a–b* and *c–d* located a half wave apart. It is understood, of course, that there are present many other lines both before and behind those represented. Also there are lines of magnetic force at right angles to the electric force. As time progresses each element of length of the line of force *a–b* moves laterally with the velocity of light. In the region where the wires are parallel, it remains straight but, upon reaching the flared section, its two ends fall behind the central section, thereby forming a curve as shown in Fig. 6.3-1(c). As this line of force moves to the end of the flared section [Fig. 6.3-1(d)], its successor *c–d* follows one-half wavelength behind.

Because of the property of inertia with which all lines of force are assumed to be endowed, the central section of a–b, which is already greatly extended due to curvature, continues in motion for some time after the two ends, attached to the conductors, have come to rest. The result is shown approximately by Fig. 6.3-1(e). An instant later and perhaps after the two ends of line of force a–b have started on their return journey, the line of force c–d approaches sufficiently close to a–b that a coalescence ensues [Fig. 6.3-1(f)]. An instant later fission takes place as illustrated in Fig. 6.3-1(g), leaving a portion of the energy of each a–b and c–d now shared by a radiated com-

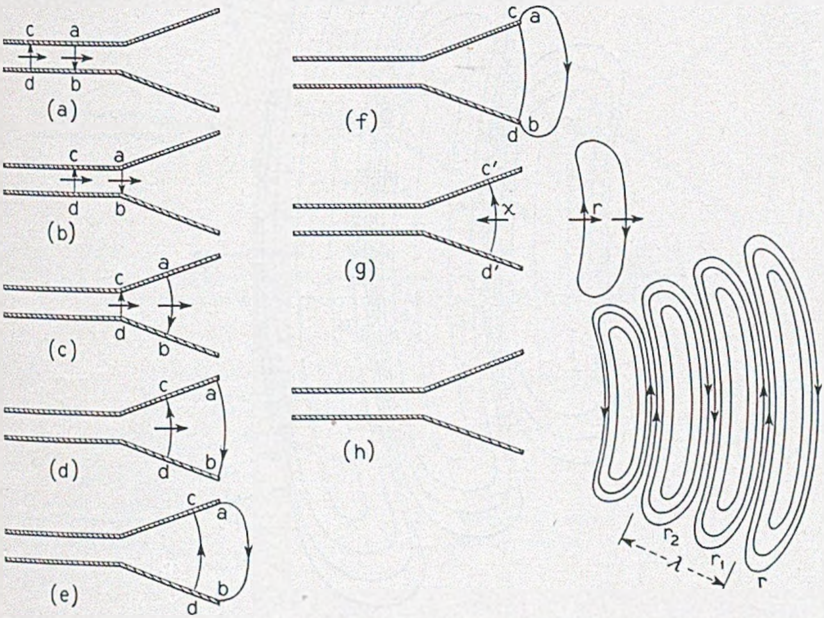


Fig. 6.3-1. Successive epochs in a highly idealized representation of radiation from the flared end of a transmission line.

ponent, r, and a reflected component, x. That the two components r and x should travel in opposite directions seems reasonable when it is noted that lines of electric force in x are in the same direction as in the adjacent portion of r. They may therefore be expected to repel. The first of these components, r, appears to the transmitter as though it were a resistance since it represents lost energy. The second, x, appears as a reactance since it represents energy returned to the transmitter. The radiated component, r, will be followed by other components r_1, r_2, etc., as represented in Fig. 6.3-1(h).

In the radiated wave front, the two components E and H are everywhere mutually perpendicular and in the same phase. Because the wave front is curved, as shown in cross section in Fig. 6.3-2, the component Poynting vectors which specify the directions in which energy is flowing will be slightly divergent. As a result, only a portion of the total wavepower will proceed in the preferred direction. It follows that, for best directivity, the emitted wave front should be substantially plane, and the lines of force should be as nearly straight as possible. There is shown in Fig. 6.3-3 a series of configura-

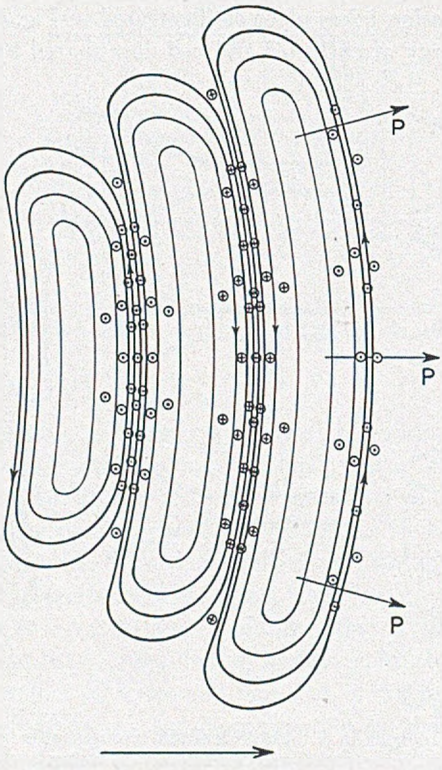


Fig. 6.3-2. Cross section of electromagnetic waves radiated from the flared end of a transmission line. Lines of electric force lie in the plane of the illustration; lines of magnetic force are perpendicular to the illustration while the flow of power is along the divergent arrows P.

tions based partly on speculation and partly on deductions from Huygens' principle. They illustrate in a rough way how, by increasing the aperture between the two wires of the elementary radiator, we may make the individual component Poynting vectors more nearly parallel.[10]

[10] Figure 6.3-3 has been greatly oversimplified. Experiment shows that, to achieve the result desired, the angle between the two wires of Fig. 6.3-3 must be smaller for larger apertures than for small apertures.

Thus far, we have restricted our considerations to directivity in the plane of the two conductors (vertical plane as here assumed). Experiment shows that, in the plane perpendicular to that illustrated, the directivity from a single pair of wires is slight. However, we may obtain additional directivity by increasing the horizontal aperture. One method of accomplishing this result is to array, at rather closely spaced intervals, identical elementary radiators each of the kind just described. [See Fig. 6.3-4(a).] An infinite number of these elements infinitesimally spaced become two parallel plates as shown in Fig. 6.3-4(b). If metal plates are now attached at the right and left

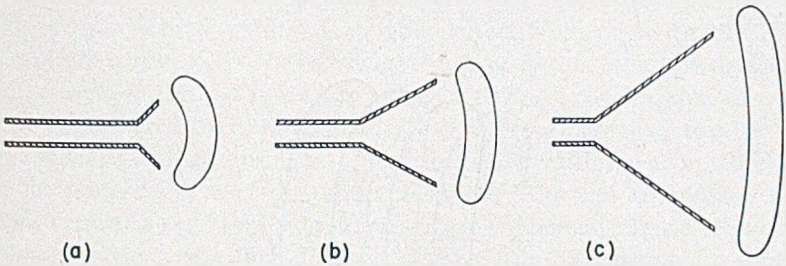


Fig. 6.3-3. Illustrating how radiating systems of large aperture may give rise to wave fronts of large radius of curvature and hence lead to increased directivity.

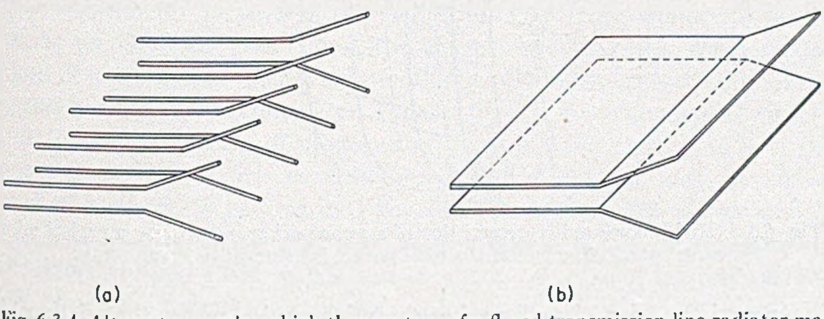


Fig. 6.3-4. Alternate ways by which the aperture of a flared transmission line radiator may be increased.

sides, the resulting configuration will become a waveguide horn. As a general rule, the larger the area of aperture, the more directive will be the antenna. The highly schematic array shown in Fig. 6.3-4(a) is introduced for illustrative purposes only. It is not one of the preferred forms used in microwave work. More practicable forms will be found in Chapter X.

The wave model shown in Fig. 6.3-2 conveys but a portion of the known facts about a radiated wave. A more accurate model is shown in skeleton form in Fig. 6.3-5. It is assumed that the transmitted wave has been launched with about equal directivity in the two principal planes and that the observer is looking into one-half of a cut-away section of the total configuration. In the complete configuration, the individual lines of electric force (solid lines) and magnetic force (dotted lines) form closed loops, thereby producing in each half-wave interval a packet of energy. The stream of projected energy from an antenna is, according to this view, a series of these packets one behind the other moving along the major axis of transmission. At the transmitter each packet may have lateral dimensions that are only slightly greater than the corresponding dimensions of the radiating antenna; but, since the packet has curvature and since propagation is radial, the packet spreads as it progresses so that at the distant receiver it may be very large indeed.

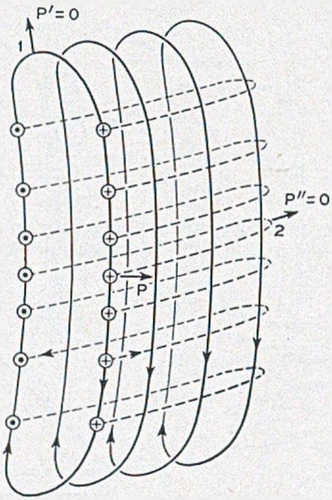


Fig. 6.3-5. Highly idealized representation of a wave-packet radiated by a typical microwave source. One half of the total packet is assumed to be cut away.

Around the edge of each packet there is a region where the relationship between the vectors E, H, and v is rather involved. For example, in the vicinity of point 1 in Fig. 6.3-5, there is a substantial component of E but at this point the vector H is zero and accordingly the Poynting vector P' at that point is also zero. (See Equation 6.1–4.) In a similar way there may be in the vicinity of point 2 a substantial component of magnetic force $\mathbf{H}$; but, since at this point the electric force is substantially zero, we conclude that the Poynting vector P'' is again zero and again no power is propagated.[11]

[11] The peculiar edge effects noted may be regarded as a result of a kind of wave interference not unlike that prevailing in the regions of minimum E and H in the case of standing waves as discussed in Section 6.3. A similar kind of wave interference is cited in Section 6.5 to account for regions of low E and H in transmission along a waveguide.

The sharpest radio beams now in general use are only a few tenths of a degree across. We conclude that for these sharp beams a small but nevertheless appreciable curvature remains in the radiated wave packet. This means that, when the wave front has arrived at a distant receiver, it is still many times larger than any receiving antenna it may be practicable to construct, and accordingly the latter can intercept but a small portion of the total advancing wavepower. This implies a considerable loss of power, which is indeed the case.

In the process of radio reception, one may think of the antenna structure as a device that cuts from the advancing wave front a segment of wavepower which it subsequently guides, preferably without reflection, to the first stages of a nearby receiver. To be efficient, the wavepower intercepted should be large. This, in turn, calls for a receiving antenna of considerable area. It will be remembered that a large aperture was also a necessary feature for high directivity at the transmitter. This is consistent with the accepted view that the processes of reception and transmission through an antenna are entirely correlative and that a good transmitting antenna is a good receiving antenna and vice versa. The directive properties of an antenna are sometimes specified in terms of its *effective area.* (See Section 10.0.)

The term *uniform plane wave* is a highly idealized entity assumed in many problems for purposes of simplicity but never quite attained in practice. In an idealized wave front, the electric and magnetic components E and H are not only everywhere mutually perpendicular but both components are exclusively transverse. That is, there is no component of either E or H in the direction of propagation. Such a wave belongs to a class known as *transverse electromagnetic waves* (TEM). These may be compared with others, to be described later, known as *transverse electric waves* (TE) and *transverse magnetic* (TM) *waves.* Waves guided along parallel conductors are also TEM waves, but except in the case of infinitely large conductors they are not *uniform plane waves.*

6.4 Reflection of Space Waves from a Metal Surface

One of the early triumphs of the electromagnetic theory was its ability to account satisfactorily for the reflection and refraction of light. This theory was so general as to include not only a wide range of wavelengths but also a wide range of surfaces as well. According to this theory, reflections may occur whenever electromagnetic waves encounter a discontinuity. This may happen, for example, when waves fall on a sheet of metal, in which case the discontinuity is due to the sudden change in conductivity. Reflection may also occur when waves are incident on a thick slab of glass or hard rubber, in which case reflection is due to a sudden change in dielectric constant.[12] Similar reflections may theoretically take place also at an interface where the permeability of the medium changes suddenly. The case in which there is a change of conductivity has an important bearing on waveguide transmission. It will therefore be discussed in considerable detail.

Assume a plane wave incident obliquely upon a conducting surface as shown in Fig. 6.4–1. The line along which the wave is progressing (wave-normal) is referred to as the *incident ray*. It intersects the conducting surface or interface at a point O and makes an angle θ with the perpendicular OZ. After reflection, the normal to the new wave wave front makes an angle θ' with the perpendicular OZ. This second wave-normal is known as the

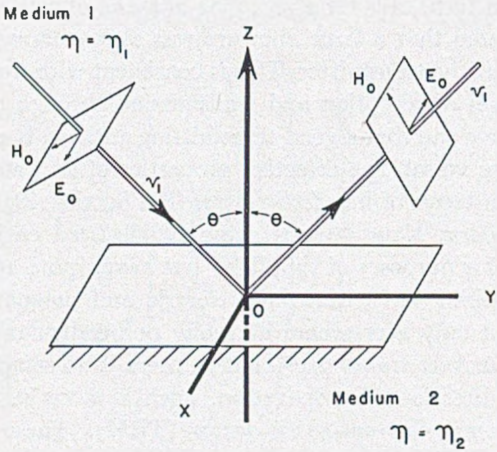


Fig. 6.4-1. Reflection at oblique incidence from a metal plate for the particular case where the electric vector is perpendicular to the plane of incidence.

reflected ray, and its angle with the perpendicular OZ is known as the *angle of reflection*. The plane containing the incident ray and the perpendicular OZ is known as the *plane of incidence*. The incident and reflected rays lie in the same plane, and their corresponding angles of incidence and reflection are numerically equal.

In problems of oblique incidence there are two cases of interest, depending on whether the electric or the magnetic component lies in the plane of incidence. For our particular purpose, the second of these two cases is of special interest and it will therefore be discussed in considerable detail. The vector relations corresponding to this case are shown in Fig. 6.4-1.

[12] For a more general discussion of the electromagnetic theory of reflection: L. Page and N. I. Adams, "Principles of Electricity," D. Van Nostrand Co., Inc., pp 569–575, New York 1931. R. I. Sarbacher and W. A. Edson, "Hyper and Ultra-high Frequency Engineering," John Wiley & Sons, Inc., pp 105–116, New York 1943.

Included are the relative directions of E and H both before and afte reflection.

In Fig. 6.4-2 there are shown in cross section representative lines of electric force in an advancing plane wave front. They are numbered respectively 1, 2, 3, 4, 5, 6, and 7. Each individual figure [(a), (b), (c), etc.] represents a succeeding period of time. We shall assume that the particular wave front singled out for illustration represents the crest of a wave Both ahead and behind this crest there are located alternately at half-wave intervals other crests and hollows, and their respective lines of force alternate in direction. Each line of force in the wave front is assumed to be moving in a direction indicated by the vector v. It is furthermore assumed that there is also present a magnetic component, indicated by the dotted vector H that is perpendicular to E and also to v. The vectors v and H must of course be so directed as to be in keeping with the right-hand or cork-screw rule, both before reflection and after reflection. Also at the point of incidence the tangential electric force must be zero. To account for this, we assume that as each line of electric force moves up to the conducting plane it is reversed in direction, thereby making on the average as many lines of electric force at the surface directed toward the observer as directed away from the observer. Consider, for example, lines of force 3 and 5, 2 and 6, and 1 and 7, in Fig. 6.4-2(c).

Associated with these two components of electric force which, let us say, are E and E', there are two components of magnetic force H and H'. These may be specified by $\mathbf{H} = \epsilon(\mathbf{v} \times \mathbf{E})$, each of which at the interface may be resolved into two components shown in Fig. 6.4-3 as $H = H_{\perp} + H_{\|}$ at the left and $H_{\perp}' = -H_{\|}'$ at the right. Combining these four vectors, assuming reflection to be perfect, we find that at the interface $H_{\perp} - H_{\perp}' = 0$ and $H_{\|} - (-H_{\|}') = 2H$, giving as an over-all result: (1) the electric force at the interface is everywhere zero; (2) the vertical component of the magnetic force at this point is also zero; and (3) the tangential component of the magnetic force at the interface is $2H$.

The peculiar configuration that resides close to the metal boundary is propagated to the right as a kind of magnetic wave. It has rather interesting properties which will become more evident by referring again to Fig. 6.4-2. Two conclusions may be drawn from this figure, depending on the point of view assumed. To a myopic observer located at the interface and unable to see far beyond the point p and unable to distinguish one line of force from another, the advancing wave front would look like a configuration of amplitude $H_{\|} = 2H$ and $E_{\|} = 0$ moving parallel to the interface with velocity $v_z = v/\sin\theta$. To this observer the apparent velocity would increase as θ becomes progressively smaller until, at perpendicular incidence, v_z would approach infinity. These results follow from the geo-

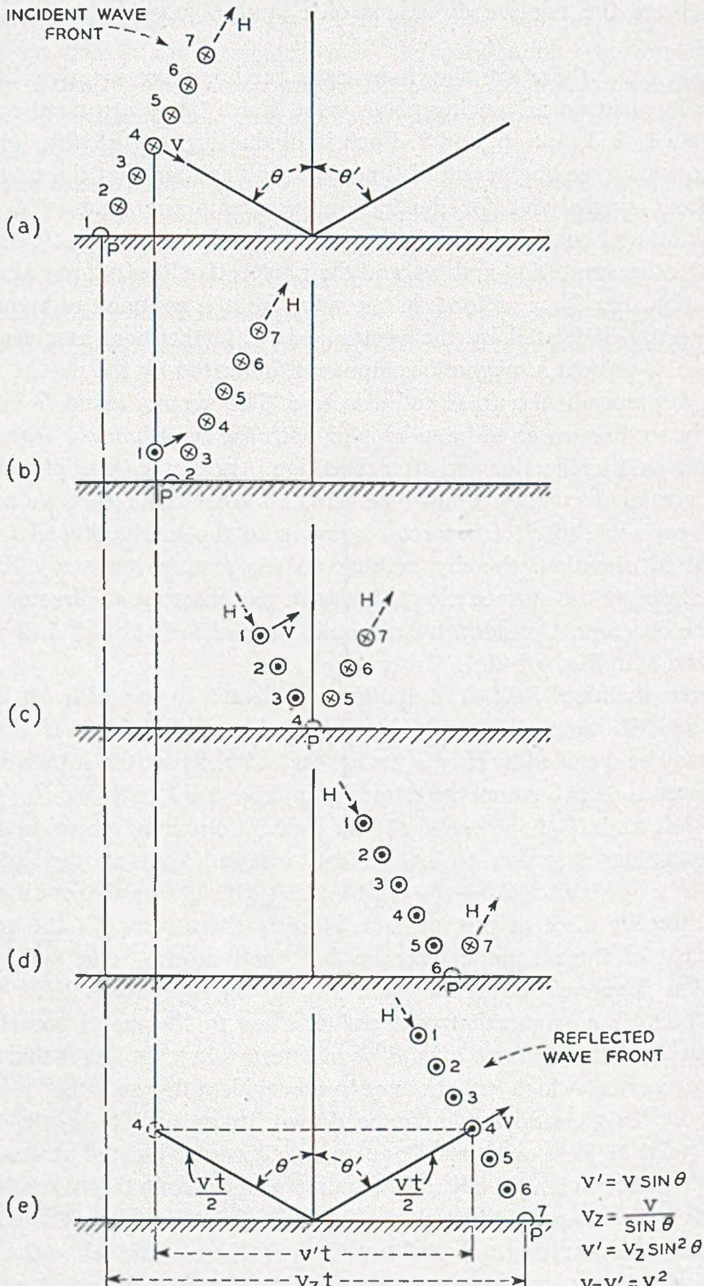


Fig. 6.4-2. Successive steps in the reflection of a single plane wave front by a metal plate.

metrical relations shown in the lower part of Fig. 6.4-2. Phenomena similar to this are sometimes observed when water waves, coming in from the ocean, break upon the beach. If the approach is nearly perpendicular, the point at which the wave breaks may proceed along the beach at a phenomenal speed. A similar effect may be produced by holding at arm's length a pair of scissors and observing the point of intersection as the blades are slowly closed. A relatively slow motion of the blades leads to a rather rapid motion of the point of intersection.

Since, in the case of incident waves, the apparent velocity is $v_z = v/\sin\theta$, the corresponding wavelength is $\lambda_z = \lambda/\sin\theta$. Both quantities play an important part in the picture of waveguide transmission to be drawn later. In particular, the apparent velocity v_z will prove to be identical with a quantity known as *phase velocity*.

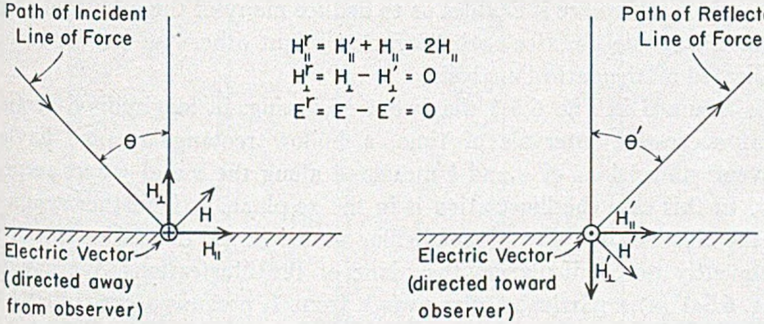


Fig. 6.4-3. Relationship between various components of E and H before and after reflection by a metal plate.

A second observer located at the interface, shown in Fig. 6.4-2, endowed with better vision and able to single out particular lines of force may obtain a somewhat different view of reflection. If he observes a particular line of force such as (4) in Fig. 6.4-2 for the considerable period of time, t, required for it to approach the conducting interface [Figs. (a) to (c)] and recede to a comparable distance [Figs. (c) to (e)], he will note that, whereas the line of force has really traveled a total distance vt, its effective progress parallel to the interface has been $v't = vt\sin\theta$. (See geometrical relations in lower part of Fig. 6.4-2.) This provides another kind of velocity ($v' = v\sin\theta$) known as *group velocity*. It is the effective velocity with which energy is propagated parallel to the metal surface. It approaches zero at perpendicular incidence. It will be observed that

$$v' = v_z \sin^2\theta$$

and

$$v'v_z = v^2 \tag{6.4-1}$$

Group velocity also plays an important part in waveguide transmission.

6.5 Waveguide Transmission

It was pointed out in an earlier chapter that each of the various configurations observed in waveguides may be considered as the resultant of a series of plane waves each traveling with a velocity characteristic of the medium inside, all multiply reflected between opposite walls. In the case of certain of these waves, this equivalence may not be readily obvious, but for the dominant mode in a rectangular guide, which is one of the more important practical cases, it is relatively simple. It also happens that the analysis of such waves throws considerable light on the nature of guided waves, and furthermore it enables us to deduce many of the useful relations used in waveguide practice—relations that might otherwise call for rather complicated mathematical analysis.

It is assumed in Fig. 6.5-1 that we are viewing, in longitudinal section and at successive intervals of time, a hollow rectangular pipe having transverse dimensions of a and b measured along the x and y axes respectively. In this case the illustration is in the xz plane. It is further assumed that the electric force lies perpendicular to the larger dimension a and is consequently perpendicular to the plane of the illustrations. We assume in Fig. 6.5-1 (a) a particular plane wave front 1, perhaps a crest, that has recently entered the guide from below. Let us say that its velocity is $v = v_a/\sqrt{\mu_r\epsilon_r}$ and that is it so directed as to make an angle θ with the left-hand wall as shown.[13] Reflection at the left-hand wall will therefore be identical with that already shown in Fig. 6.4-2. A portion of the wave front that has just previously undergone reflection is shown immediately below at 2 in Fig. 6.5-1(a). We assume further that this front is made up of lines of electric force perpendicular to the illustration together with associated lines of magnetic force lying in the plane of the illustration. It will be obvious presently that, like the case of reflection from a single conducting sheet discussed in the previous section, we may obtain two rather different pictures of what takes place within the guide, depending on whether we fix our attention on the configuration as a whole or on some particular line of force which we may identify and follow through a considerable interval of time. We shall first consider the configuration as a whole.

[13] It is to be noted that the angle ϕ which the wave front makes with the metal wall is equal to the angle which the wave-normal (ray) makes with the perpendicular to the metal wall.

We show in Fig. 6.5-1(b) the same wave front shown in Fig. 6.5-1(a) but at an epoch later—after it has progressed a considerable distance along the guide. We now find the reflected portion 2 complete and a new portion

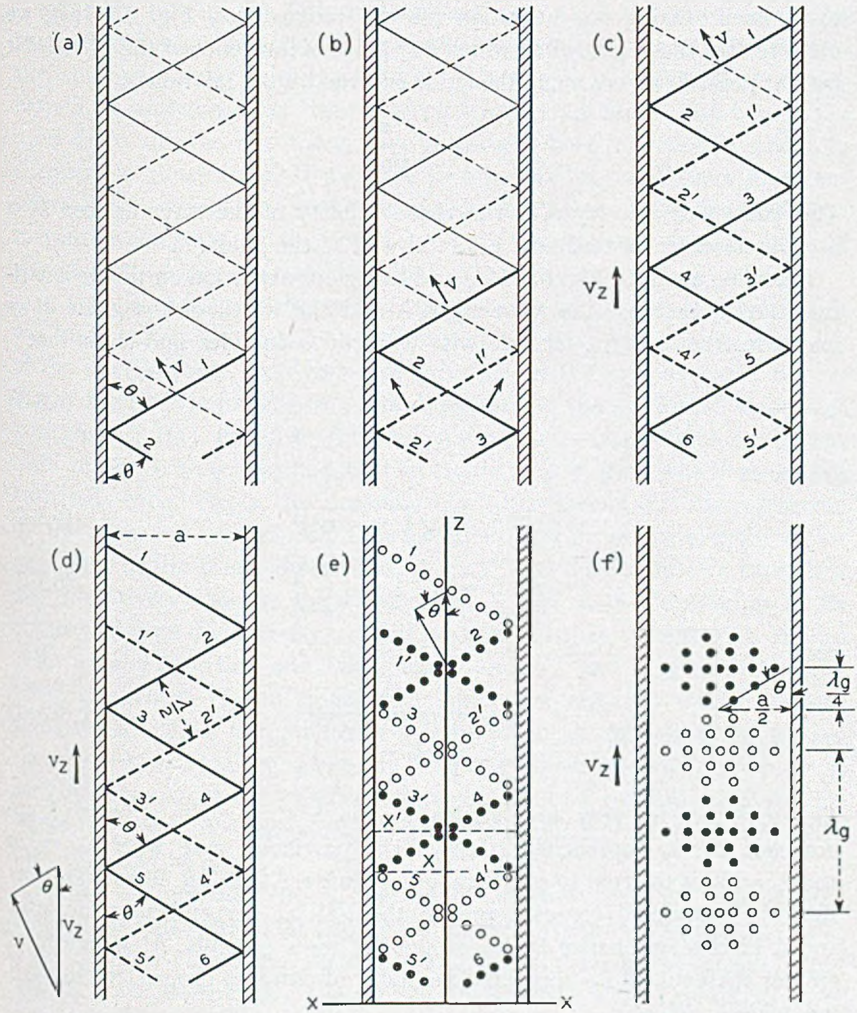


Fig. 6.5-1. The propagation of a multiply reflected wave front between two metal plates [Figs. (a)–(d)] is equivalent to the transmission of a TE wave parallel to the two plates. [Fig. (f)].

3 about to enter the guide. Following wave front 1 and at a distance of one-half wave behind, we find, shown dotted, the "hollow" of the wave. This we shall designate by the numeral 1′. We find here also a new portion of the "hollow" 2′ that has just undergone reflection.

In Fig. 6.5-1(c) and again in Fig. 6.5-1(d) we find successive positions of these same wave fronts as they have moved forward in the guide. We may, if we like, think of these fronts as discrete waves moving zig-zag through the guide or as a single large wave front folded repeatedly back upon itself. Fixing our attention for the moment on Fig. 6.5-1(d), we observe that the velocity v at which any point of incidence of the wave front (say at point 5) moves along the guide is given by the relation

$$v_z = \frac{v}{\sin \theta}$$

This particular velocity v_z is the phase velocity of the wave as seen by a myopic observer located near a lateral wall of the guide.

Referring again to Fig. 6.5-1(d) and fixing our attention on the geometrical relation between the wavelength λ and the width of the guide a, we may construct a right triangle with $\lambda/2$ and a as sides and show that

$$\cos \theta = \frac{\lambda}{2a} \tag{6.5–1}$$

and since

$$\sin \theta = \sqrt{1 - \cos^2 \theta} \tag{6.5–2}$$

$$\sin \theta = \sqrt{1 - \left(\frac{\lambda}{2a}\right)^2} \tag{6.5–3}$$

and

$$v_z = \frac{v}{\sqrt{1 - \left(\frac{\lambda}{2a}\right)^2}} \tag{6.5–4}$$

This says that for very large guides, that is, $\lambda < 2a$, $v_z \doteq v$, but as λ approaches $2a$, v_z approaches infinity. The particular case where $\lambda = 2a$ and $v_z = \infty$ is referred to as the *cut-off condition*. At cut-off, it would appear that the individual waves approach the wall at perpendicular incidence and a kind of resonance between opposite walls prevails. At wavelengths greater than cut-off no appreciable amount of power is propagated through the guide.

The particular value of wavelength measured in air, corresponding to cut-off, is referred to as the *critical* or *cut-off wavelength* and is designated thus: $\lambda_c = 2a$. The corresponding frequency is similarly known as the *critical* or *cut-off frequency* and it is designated thus: $f_c = v/\lambda_c$. It is sometimes convenient to designate the ratio of the operating wavelength to the critical wavelength by the symbol ν. From Equation 6.5–4 it follows that

$$\frac{v_z}{v} = \frac{1}{\sqrt{1 - \left(\frac{\lambda}{\lambda_c}\right)^2}} = \frac{1}{\sqrt{1 - \left(\frac{f_c}{f}\right)^2}} = \frac{1}{\sqrt{1 - \nu^2}} \qquad (6.5\text{–}5)$$

Referring to Fig. 6.5-1(a) we have indicated that the wave front 1 is made up of lines of electric force directed through the plane of the illustration and hence away from the observer. There are, of course, lines of magnetic force and also other lines of electric force both ahead and behind the wave front drawn, but these have purposely been omitted in order to simplify the illustration. If we were to take the magnetic force into consideration we would find as in Fig. 6.4-2 that, at the reflecting surface, a tangential component only is present and its magnitude is twice that of the magnetic component of the incident wave.

In the discussion of reflection of plane waves in the previous section, it was also pointed out that the act of reflecting a wave reverses the direction of the electric force. Applying this principle to the case at hand, we see that if the electric force is directed downward in the section of wavefront 1 of Fig. 6.5-1(a), it will be directed upward in 2. Carrying this idea forward to Fig. 6.5-1(e) we find that in fronts 1, 2, 3, etc., which we rather arbitrarily called crests, the electric vector alternates in direction as shown by the open and solid circles. Likewise the direction of the electric vector alternates in the fronts designated as 1′, 2′, and 3′, but in this case they are respectively opposite in direction to 1, 2, and 3. Continuing to fix our attention on Fig. 6.5-1(e), it will be observed that the direction of lines of force is the same in 1′ and 2, in 2′ and 3, and in 3′ and 4, indefinitely along the entire length of the guide. Thus there are regularly spaced regions along the length of the guide where the electric vector is directed toward the observer alternating with other regions where the electric vector is directed away from the observer. Between the two are still other regions where the respective component vectors are oppositely directed and hence their sum may be zero.

Adding the foregoing effects, bearing in mind that there are lines of force both ahead and behind the highly simplified wave fronts shown, we have a new wave configuration moving parallel to the main axis of the guide with a phase velocity v_z as suggested by Fig. 6.5-1(f). Examining more carefully the wave interference that is here taking place, it becomes evident that if we pass laterally across the guide along the line x in Fig. 6.5-1(e) the instantaneous value of the resultant electric vector as shown is everywhere zero. On the other hand, if we cross the guide along a parallel line x', the electric vector varies sinusoidally beginning at zero at either wall and reaching a maximum in the middle of the guide. It will be observed that if we pass along the major axis z of the guide the electric vector at any instant again varies sinusoidally with distance. However, at the boundary of the guide the resultant electric vector is everywhere zero. Since there was no component of the electric force lying along the axis z of the guide in the component waves that gave rise to this configuration, there can be no such component in the resultant. Waves in which the electric vector is exclusively transverse are known as *transverse electric*, or TE, *waves*.

A complete account of transmission of this kind should include, of course, a consideration of the lines of magnetic force. From Fig. 6.4-3 it is evident that, at the point of reflection of the component plane wave on the guide wall, there are two components of magnetic force $H_{\perp}$ and $H_{\parallel}$ in both the incident and reflected waves. When these are added, the resultant of the transverse magnetic force, like that of the electric force, differs at different points in the guides. Following alone the line x', it is found that for the particular condition here assumed, the magnetic force is zero at each wall increasing sinusoidally to a maximum midway between. At this point the magnetic component is entirely transverse. Following along the line x, it will be found that the magnetic vector is a maximum near each wall decreasing cosinusoidally to zero in the middle. It is of particular interest that, at the wall of the guide, the magnetic component lies parallel to the axis. Magnetic lines of force are, in this type of wave, closed loops, whereas lines of electric force merely extend from the upper to the lower walls of the guide. The arrangement of lines of electric and magnetic force in this type of wave is shown in Fig. 5.2-1. The quantitative relationships between the various components of E and H are specified more definitely by Equation 5.2-1. The significance of the wavelength λ_g of this new configuration will be obvious from Fig. 6.5-1(f).

There are certain useful results that follow from Fig. 6.5-1(f). It may be seen from the triangle there shown that

$$\frac{\lambda_g}{4} = \frac{a}{2} \cot \theta \qquad (6.5\text{-}6)$$

From Equations 6.5-1 and 6.5-3, it will also be seen that

$$\cot \theta = \frac{\cos \theta}{\sin \theta} = \frac{\lambda}{2a\sqrt{1 - \left(\frac{\lambda}{2a}\right)^2}} \qquad (6.5\text{-}7)$$

Therefore

$$\lambda_g = \frac{\lambda}{\sqrt{1 - \left(\frac{\lambda}{2a}\right)^2}} = \frac{\lambda}{\sqrt{1 - \nu^2}} \qquad (6.5\text{-}8)$$

Since $1/\sqrt{1 - \nu^2}$ is the ratio of the apparent wavelength in the guide to that in free space and since for hollow pipes it is greater than unity, it is sometimes referred to as the *stretching factor*. It appears frequently in quantitative expressions relating to waveguides. Since velocity is equal to the number of waves passing per second times the length of each wave, we have

$$v_z = \frac{v}{\sqrt{1 - \nu^2}} \tag{6.5-9}$$

This is equivalent to the relation shown as Equation 6.5-5.

A matter of special interest is the rate at which energy is propagated along the guide. For present purposes, it is convenient to regard a moving line of force and its associated magnetic force as a unit of propagated energy. A knowledge of the path followed by such a line of force will therefore shed light on the rate at which energy is propagated along a waveguide.

It was pointed out in connection with Equation 6.4-2 that, when a wave is incident obliquely upon a metal surface, the apparent phase of the wave progresses at a velocity v_z greater than the velocity of light v, but that the energy actually progresses parallel to the interface at a velocity v' less than the velocity of light. It was pointed out, too, that $v' = v \sin \theta = v_z \sin^2 \theta$. Because of multiple reflections between opposite walls of a waveguide, its *phase velocity* is identical with v_z. Also, because of these multiple reflections, energy being carried by these component plane waves follows a rather devious zig-zag path and will therefore progress along the axis of the guide at a relatively slow rate. This velocity which is known as the *group velocity* is idential with v' above. From relations already given, it will be seen that

$$v' = v\sqrt{1 - \nu^2} \tag{6.5-10}$$

also

$$v' = v_z(1 - \nu^2) \tag{6.5-11}$$

It will be apparent from this relation that, at cut-off, where $\nu = 1$, energy is propagated along the guide with zero velocity. This is consistent with the idea already set forth that, at cut-off, energy oscillates back and forth between opposite faces of the guide. As we leave cut-off and progress toward higher frequencies (shorter waves), the group velocity v' increases as the phase velocity v_z decreases, until, at extremely high frequencies, both approach the velocity v characteristic of the medium. This relationship is made more evident by Fig. 6.5-2.

Reviewing again the simple analysis just made, we find that the wave configuration that actually progresses along a conventional rectangular waveguide may be regarded as the result of interference of ordinary uniform plane waves multiply reflected between opposite walls of the guide. This viewpoint accounts for not only the distribution of the lines of force in the wave front but also for the velocity at which the phase progresses and the velocity at which energy is propagated. As we shall soon see, it accounts also for the rate of attenuation.

In the particular configuration just described the electric component is everywhere transverse, whereas the magnetic component may be either longitudinal or transverse, depending on the point in a guide at which observations are made. These waves are plane waves, but, since the elec-

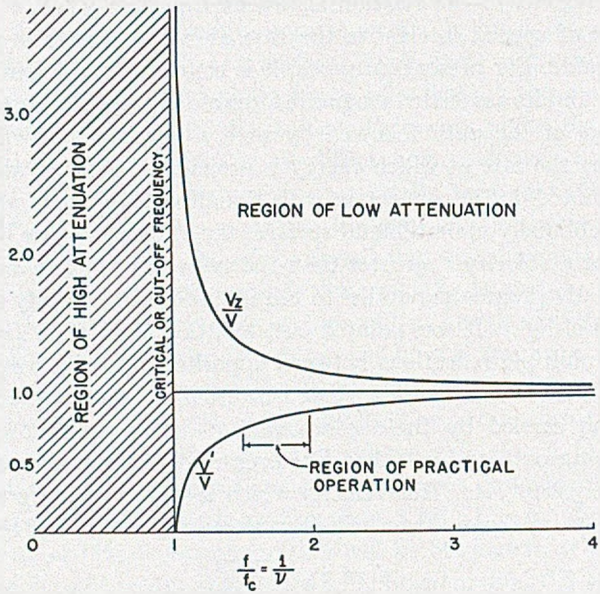


Fig. 6.5-2. Relative phase velocity v_z and group velocity v' for various conditions of operation of a waveguide.

tric intensity is not uniformly distributed over the wave front, they are not uniform plane waves.

The concept of multiply reflected waves provides a basis for calculating the attenuation in rectangular guides as was shown by John Kemp several years ago.[11] The procedure is outlined briefly below. The reader is referred to the published article for details.

There is shown in Fig. 6.5-3 a short section of hollow waveguide in which we imagine multiply reflected plane waves are propagated. We fix our attention on a zig-zag section cut from the guide and so directed that it

[11] John Kemp, "Electromagnetic Waves in Metal Tubes of Rectangular Cross-section," *Jour. I.E.E.*, Part III, Vol. 88, No. 3, pp 213–218, September 1941.

lies parallel to the direction of propagation of the elemental wave fronts. The top and bottom conductors so formed may be regarded as a uniform flat-conductor transmission line with oblique reflecting plates (sections of the side walls) spaced at regular intervals. Other transmission lines adjacent to that under consideration behave in exactly the same way as that singled out for examination and at the same time act as guard plates to insure that the lines of force so propagated remain straight.

It is clear that the attenuation in each elemental transmission line will be that incidental to losses in the upper and lower conductors plus the losses incidental to reflection at oblique incidence from the several reflecting

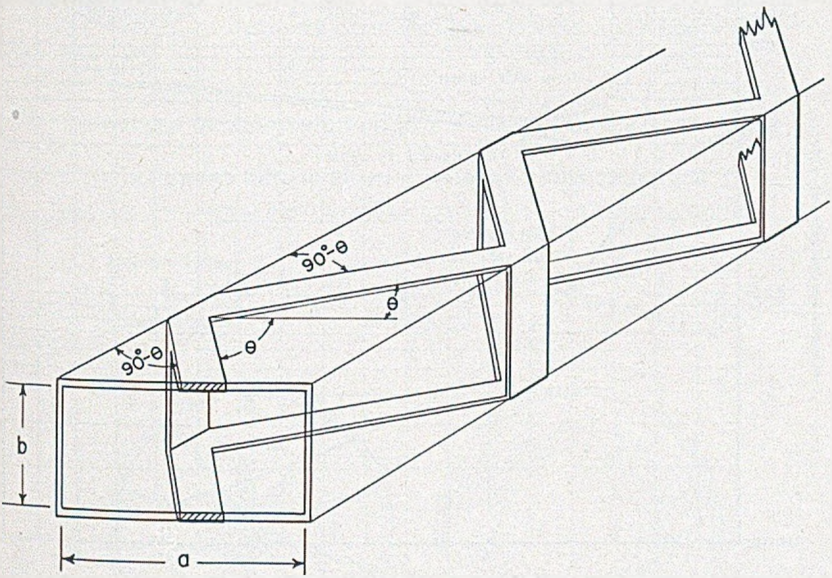


Fig. 6.5-3. Elementary transmission lines terminated periodically by reflecting plates which go to make up a rectangular waveguide.

plates. The total attenuation of the rectangular guide may then be found by summing up over a unit length of waveguide all of the elemental lines. This has been done with results that are equivalent to the corresponding equations given in Chapter V. The results are plotted in Fig. 6.5-4.

Certain characteristics of these curves may be readily accounted for. For instance, at cut-off ($\theta = 0$), both the number of unit reflection plates and the number of flat-plate transmission lines in a given length of waveguide will be infinite. As a result, the component attenuations arising in each of these two sources will likewise be infinite. As the frequency is increased above cut-off the angle θ will increase accordingly, leading thereby to fewer side-wall reflections and to a shorter over-all length of zig-zag transmission line. Thus, in this frequency range, the attenuations contributed both by the side walls and by the top and bottom plates decrease with increasing frequency. Proceeding to frequencies far above cut-off, where θ approaches 90 degrees, there will not only be very few reflections but the over-all length of zig-zag line will approach as its limit a single, straight two-conductor line made up of the top and bottom plates alone. Thus the attenuation due to the side walls will approach zero and that due to the top and bottom plates will increase as the square root of the

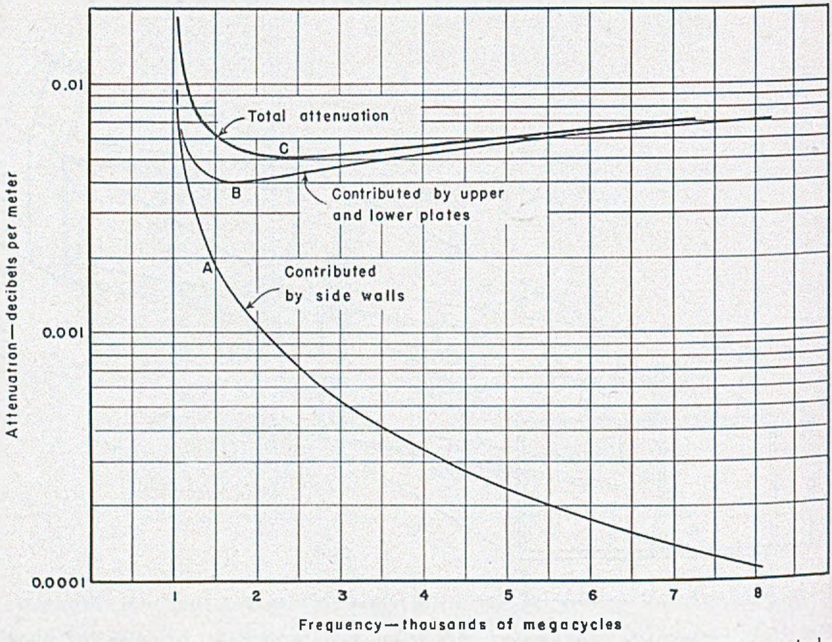


Fig. 6.5-4. Component attentuations contributed by the top and bottom plates and also the two side walls of a rectangular waveguide.

frequency. Since the attenuation contributed by the top and bottom plates first decreases but later increases with frequency, we may expect, between these two ranges, a region of minimum attenuation. The attenuations contributed by the upper and lower plates and also by the side walls of a 7.5 cm × 15 cm copper guide carrying the dominant mode have been calculated. The results have been plotted as curves A and B in Fig. 6.5-4. They follow the courses predicted by the preceding qualitative reasoning.

The fact that the reflection type of attenuation, such as is evident in the side walls above, decreases with frequency, suggests that, if a kind of waveguide could be devised where this type of attenuation alone exists, we could then operate the guide at extremely high frequencies and thereby obtain relatively low attenuations. This can, in effect, be done. It calls for a guide of circular cross section and a special configuration, known as

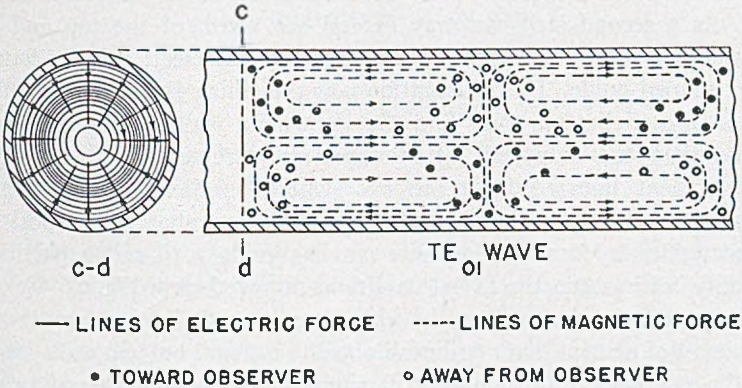


Fig. 6.5-5. The circular electric or TE_{01} configuration in a circular waveguide.

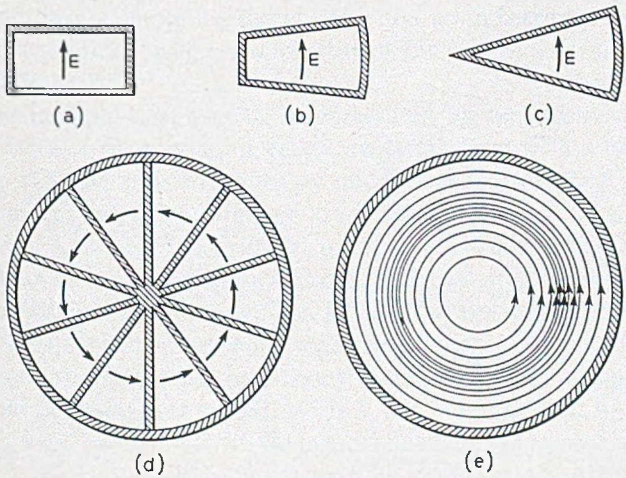


Fig. 6.5-6. Evolution of the circular-electric wave in a circular pipe from a dominant wave in a rectangular pipe.

the *circular-electric wave.* In this configuration, the resultant electric force is everywhere parallel to the conducting boundary as shown in Fig. 6.5-5.

That such a wave will lead to the interesting frequency characteristic noted is made more plausible by referring to Fig. 6.5-6 and its associated discussion. Figure 6.5-6(a) shows a conventional form of rectangular guide in which plane waves are multiply reflected from the two short sides.

In Fig. 6.5-6(b) the proportions of the guide have been altered somewhat, but since the lines of electric force are still perpendicular to the top and bottom plates, the guide may be expected to function substantially as before. At the most, some attenuation that previously originated in the left-hand side wall may now be transferred to the top and bottom walls. As a second step, we may extend the width of the top and bottom walls as shown in Fig. 6.5-6(c) until they intersect, thereby forming an arc-shaped guide. The attenuation now prevailing is evidently confined to the top and bottom walls and the right-hand wall. It is reasonable to assume that the side wall attenuation still decreases with frequency since incident lines of force are everywhere parallel to this wall. As a third step, we assemble as in Fig. 6.5-6(d) a number of identical arc-shaped guides to form a composite circular guide with radial partitions. If, finally, we imagine the radial partitions removed as in Fig. 6.5-6(e), the resulting configuration will not be altered and we shall have removed the component of attenuation attributable to the top and bottom walls leaving only the component of attenuation attributable to the one side wall, which, as we have pointed out, becomes progressively smaller as the frequency is indefinitely increased.

Memory Requirements in a Telephone Exchange

By CLAUDE E. SHANNON

(*Manuscript Received Dec. 7, 1949*)

1. Introduction

A GENERAL telephone exchange with N subscribers is indicated schematically in Fig. 1. The basic function of an exchange is that of setting up a connection between any pair of subscribers. In operation the exchange must "remember," in some form, which subscribers are connected together until the corresponding calls are completed. This requires a certain amount of internal memory, depending on the number of subscribers, the maximum calling rate, etc. A number of relations will be derived based on these considerations which give the minimum possible number of relays, crossbar switches or other elements necessary to perform this memory function. Comparison of any proposed design with the minimum requirements obtained from the relations gives a measure of the efficiency in memory utilization of the design.

Memory in a physical system is represented by the existence of stable internal states of the system. A relay can be supplied with a holding connection so that the armature will stay in either the operated or unoperated positions indefinitely, depending on its initial position. It has, then, two stable states. A set of N relays has 2^N possible sets of positions for the armatures and can be connected in such a way that these are all stable. The total number of states might be used as a measure of the memory in a system, but it is more convenient to work with the logarithm of this number. The chief reason for this is that the amount of memory is then proportional to the number of elements involved. With N relays the amount of memory is then $M = \log 2^N = N \log 2$. If the logarithmic base is two, then $\log_2 2 = 1$ and $M = N$. The resulting units may be called binary digits, or more shortly, bits. A device with M bits of memory can retain M different "yes's" or "no's" or M different 0's or 1's. The logarithmic base 10 is also useful in some cases. The resulting units of memory will then be called decimal digits. A relay has a memory capacity of .301 decimal digits. A 10×10 crossbar switch has 100 points. If each of these points could be operated independently of the others, the total memory capacity would be 100 bits or 30.1 decimal digits. As ordinarily used, however, only one point in a vertical can be closed. With this restriction the capacity is one decimal digit for each vertical, or a total of ten decimal digits. The panels used in a panel type exchange are another form of memory device. If the commutator in a panel has 500 possible levels, it has a memory capacity of log 500;8.97 bits or 2.7 decimal digits. Finally, in a step-by-step system, 100-point selector switches are used. These have a memory of two decimal digits.

Frequently the actual available memory in a group of relays or other devices is less than the sum of the individual memories because of artificial restrictions on the available states. For technical reasons, certain states are made inaccessible—if relay A is operated relay B must be unoperated, etc. In a crossbar it is not desirable to have more than nine points in the same horizontal operated because of the spring loading on the crossarm. Constraints of this type reduce the memory per element and imply that more than the minimum requirements to be derived will be necessary.

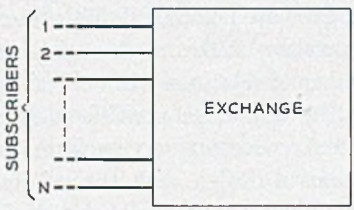


Fig. 1—General telephone exchange.

2. Memory Required for any S Calls out of N Subscribers

The simplest case occurs if we assume an isolated exchange (no trunks to other exchanges) and suppose it should be able to accommodate any possible set of S or fewer calls between pairs of subscribers. If there are a total of N subscribers, the number of ways we can select m pairs is given by

$$\frac{N(N-1)(N-2)\cdots(N-2m+1)}{2^m m!} = \frac{N!}{2^m m!(N-2m)!} \tag{1}$$

The numerator $N(N-1)\cdots(N-2m+1)$ is the number of ways of choosing the $2m$ subscribers involved out of the N. The $m!$ takes care of the permutations in order of the calls and 2^m the inversions of subscribers in pairs. The total number of possibilities is then the sum of this for $m = 0, 1, \cdots, S$; i.e.

$$\sum_{m=0}^{S} \frac{N!}{2^m m!(N-2m)!} \tag{2}$$

The exchange must have a stable internal state corresponding to each of these possibilities and must have, therefore, a memory capacity M where

$$M = \log \sum_{0}^{S} \frac{N!}{2^m m!(N-2m)!}. \tag{3}$$

If the exchange were constructed using only relays it must contain at least $\log_2 \sum N!/2^m m!(N - 2m)!$ relays. If 10×10 point crossbars are used in the normal fashion it must contain at least $\frac{1}{10} \log_{10} \sum N!/2^m m!(N - 2m)!$ of these, etc. If fewer are used there are not enough stable configurations of connections available to distinguish all the possible desired interconnections. With $N = 10{,}000$, and a peak load of say 1000 simultaneous conversations $M = 16{,}637$ bits, and at least this many relays or 502 10×10 crossbars would be necessary. Incidentally, for numbers N and S of this magnitude only the term $m = S$ is significant in (3).

The memory computed above is that required only for the basic function of remembering who is talking to whom until the conversation is completed. Supervision and control functions have been ignored. One particular supervisory function is easily taken into account. The call should be charged to

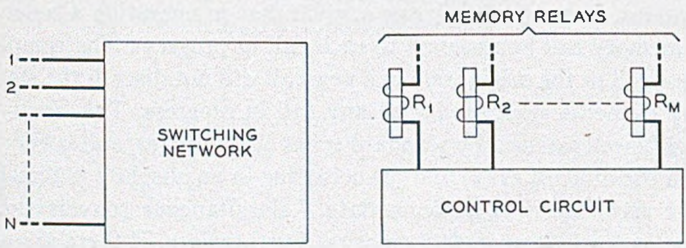


Fig. 2—Minimum memory exchange.

the calling party and under his control (i.e. the connection is broken when the calling party hangs up). Thus the exchange must distinguish between a calling b and b calling a. Rather than count the number of pairs possible we should count the number of ordered pairs. The effect of this is merely to eliminate the 2^m in the above formulas.

The question arises as to whether these limits are the best possible—could we design an exchange using only this minimal number of relays, for example? The answer is that such a design is possible in principle, but for various reasons quite impractical with ordinary types of relays or switching elements. Figure 2 indicates schematically such an exchange. There are M memory relays numbered $1, 2, \ldots, M$. Each possible configuration of calls is given a binary number from 0 to 2^M and associated with the corresponding configuration of the relay positions. We have just enough such positions to accommodate all desired interconnections of subscribers.

The switching network is a network of contacts on the memory relays such that when they are in a particular position the correct lines are connected together according to the correspondence decided upon. The control circuit is essentially merely a function table and requires, therefore, no memory. When a call is completed or a new call originated the desired configuration of the holding relays is compared with the present configuration and voltages applied to or eliminated from all relays that should be changed.

Needless to say, an exchange of this type, although using the minimum memory, has many disadvantages, as often occurs when we minimize a design for one parameter without regard to other important characteristics. In particular in Fig. 2 the following may be noted: (1) Each of the memory relays must carry an enormous number of contacts. (2) At each new call or completion of an old call a large fraction of the memory relays must change position, resulting in short relay life and interfering transients in the conversations. (3) Failure of one of the memory relays would put the exchange completely out of commission.

3. The Separate Memory Condition

The impracticality of an exchange with the absolute minimum memory suggests that we investigate the memory requirements with more realistic assumptions. In particular, let us assume that in operation a separate part of the memory can be assigned to each call in progress. The completion of a current call or the origination of a new call will not disturb the state of the memory elements associated with any call in progress. This assumption is reasonably well satisfied by standard types of exchanges, and is very natural to avoid the difficulties (2) and (3) occurring in an absolute minimal design.

If the exchange is to accommodate S simultaneous conversations there must be at least S separate memories. Furthermore, if there are only this number, each[1] of these must have a capacity $\log \frac{N(N-1)}{2}$. To see this, suppose all other calls are completed except the one in a particular memory. The state of the entire exchange is then specified by the state of this particular memory. The call registered here can be between any pair of the N subscribers, giving a total of $N(N-1)/2$ possibilities. Each of these must correspond to a different state of the particular memory under consideration, and hence it has a capacity of least $\log N(N-1)/2$.

The total memory required is then

$$M = S \log \frac{N(N-1)}{2}. \tag{4}$$

If the exchange must remember which subscriber of a pair originated the call we obtain

$$M = S \log N(N-1). \tag{5}$$

or, very closely when N is large,

$$M = 2S \log N. \tag{6}$$

[1] B. D. Holbrook has pointed out that by using more than S memories, each can have for certain ratios of $\frac{S}{N}$, a smaller memory, resulting in a net saving. This only occurs, however, with unrealistically high calling rates.

The approximation in replacing (5) by (6), of the order of $\frac{S}{N} \log e$, is equivalent to the memory required to allow connections to be set up from a subscriber to himself. With $N = 10{,}000$, $S = 1{,}000$, we obtain $M = 26{,}600$

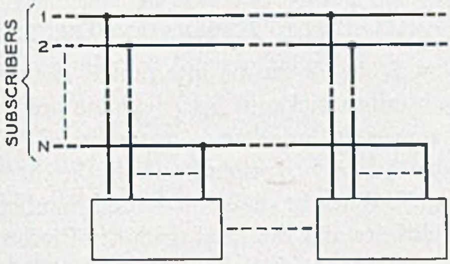


Fig. 3—Minimum separate memory exchange.

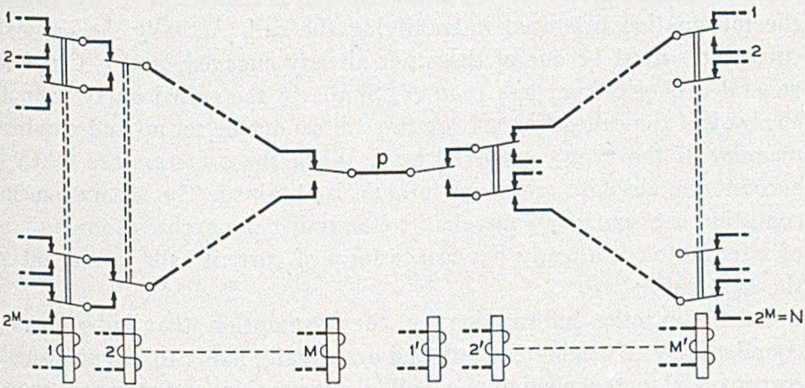


Fig. 4—Interconnecting network for Fig. 3.

from (6). The considerable discrepancy between this minimum required memory and the amount actually used in standard exchanges is due in part to the many control and supervision functions which we have ignored, and in part to statistical margins provided because of the limited access property.

The lower bound given by (6) is essentially realized with the schematic exchange of Fig. 3. Each box contains a memory $2 \log N$ and a contact network capable of interconnecting any pair of inputs, an ordered pair being associated with each possible state of the memory. Figure 4 shows such an interconnection network. By proper excitation of the memory relays 1, 2, $\cdots$, M, the point p can be connected to any of the $N = 2^m$ subscribers on the left. The relays $1'$, $2'$, $\cdots$, M' connect p to the called subscriber on the right. The general scheme of Fig. 3 is not too far from standard methods, although the contact load on the memory elements is still impractical. In actual panel, crossbar and step-by-step systems the equivalents of the memory boxes are given limited access to the lines in order to reduce the contact loads. This reduces the flexibility of interconnection, but only by a small amount on a statistical basis.

4. Relation to Information Theory

The formula $M = 2S \log N$ can be interpreted in terms of information theory.[2] When a subscriber picks up his telephone preparatory to making a call, he in effect singles out one line from the set of N, and if we regard all subscribers as equally likely to originate a call, the corresponding amount of information is $\log N$. When he dials the desired number there is a second choice from N possibilities and the total amount of information associated with the origin and destination of the call is $2 \log N$. With S possible simultaneous calls the exchange must remember $2S \log N$ units of information.

The reason we obtain the "separate memory" formula rather than the absolute minimum memory by this argument is that we have overestimated the information produced in specifying the call. Actually the originating subscribers must be one of those not already engaged, and is therefore in general a choice from less than N. Similarly the called party cannot be engaged; if the called line is busy the call cannot be set up and requires no memory of the type considered here. When these factors are taken into account the absolute minimum formula is obtained. The separate memory condition is essentially equivalent to assuming the exchange makes no use of information it already has in the form of current calls in remembering the next call.

Calculating the information on the assumption that subscribers are equally likely to originate a call, and are equally likely to call any number, corresponds to the maximum possible information or "entropy" in communication theory. If we assume instead, as is actually the case, that certain interconnections have a high *a priori* probability, with others relatively small, it is possible to make a certain statistical saving in memory.

This possibility is already exploited to a limited extent. Suppose we have two nearby communities. If a call originates in either community, the probability that the called subscriber will be in the same community is much greater than that of his being in the other. Thus, each of the exchanges can be designed to service its local traffic and a small number of intercommunity calls. This results in a saving of memory. If each exchange has N subscribers and we consider, as a limiting case, no traffic between exchanges,

[2] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, Vol. 27, pp. 379–423, and 623–656, July and October 1948.

the total memory by (6) would be $4S \log N$, while with all $2N$ subscribers in the same exchange $4S \log 2N$ would be required.

The saving just discussed is possible because of a group effect. There are also statistics involving the calling habits of individual subscribers. A typical subscriber may make ninety per cent of his calls to a particular small number of individuals with the remaining ten per cent perhaps distributed randomly among the other subscribers. This effect can also be used to reduce memory requirements, although paper designs incorporating this feature appear too complicated to be practical.

Acknowledgment

The writer is indebted to C. A. Lovell and B. D. Holbrook for some suggestions incorporated in the paper.

Matter, A Mode of Motion

By R. V. L. HARTLEY

(*Manuscript Received Feb. 28, 1950*)

Both the relativistic and wave mechanical properties of particles appear to be consistent with a picture in which particles are represented by localized oscillatory disturbances in a mechanical ether of the MacCullagh-Kelvin type. Gyrostatic forces impart to such a medium an elasticity to rotation, such that, for very small velocities, its approximate equations are identical with those of Maxwell for free space. The important results, however, follow from the inherent non-linearity of the complete equations and the time dependence of the elasticity associated with finite displacements. These lead to reflections which permit of a wave of finite energy remaining localized. Because of the non-linearity, the amplitude and energy of a stable mode, as well as the frequency, are determined by the constants of the medium. Such a stable mode is capable of translational motion and so is suitable to represent a particle. The mass assigned to it is derived from its energy by the relativity relation. While this mass is dimensionally the same as that of the medium it is differently related to the energy and so need not conform to the classical laws which the latter is assumed to obey.

Exchanges of energy between particles and between a particle and radiation involve frequency changes as in the quantum theory. The experimental detection of a uniform velocity relative to the medium is not to be expected. Besides providing a new approach to the problems of particle mechanics, the theory offers the prospect of incorporating the present pictures into a more comprehensive one, with a material reduction in the number and complexity of the independent assumptions.

Introduction

THE following quotation states a conclusion which is widely held: "But in view of the more recent development of electrodynamics and optics it became more and more evident that classical mechanics affords an insufficient foundation for the physical description of all natural phenomena."[1] This implies that classical mechanics and classical electromagnetics are so alike that one may be condemned for the shortcomings of the other. Actually, classical electromagnetics is in open disagreement with classical mechanics particularly with respect to those features for which it has been most criticized. According to the mechanical principle of relativity,[2] the equations of any mechanical system are invariant under the Newtonian transformation, $x = x' + Vt', y = y', z = z', t = t'$, where V is a constant velocity in the x direction. Since the classical electromagnetic equations are not invariant under this transformation, they cannot describe the performance of any classical mechanical system. Their failures, therefore, should not stand in the way of a study of the possibilities of such systems.

The system considered here is the so-called rotational ether, suggested

[1] A. Einstein, The Theory of Relativity, Methuen & Co., Ltd., London, 1921, p. 13.
[2] Haas, Introduction to Theoretical Physics, 2nd Ed., Vol. I, p. 46.

by MacCullagh and elaborated by Kelvin, in which the stiffness is associated with gyrostatic forces. Some consideration has been given to an alternative model consisting of a non-viscous liquid in a high state of fine scale turbulence. It is well known that, by virtue of the gyrostatic forces associated with it, a vortex will transmit a wave of transverse displacement along its axis. It would appear, therefore, that a gross wave involving similar displacements would be passed along from vortex to vortex, much as a sound wave is passed from molecule to molecule. However, since this model has not yet been shown to be fully equivalent to Kelvin's, attention will be confined to the latter. While this, as developed by Kelvin, gave a satisfactory description of electromagnetic waves in free space, it had nothing to represent matter. This was assumed to be something different from ether, which might or might not be pervaded by it. A closer study of the model has indicated that the peculiar nature of its stiffness makes possible sustained oscillatory disturbances in which the energy remains localized about a center which may move with any velocity less than that of a free wave. It is proposed to use such quasi-standing wave patterns to describe material particles. Matter, then, has no existence apart from the ether, and the motion of particles is the motion of patterns of mechanical wave motion. While the ether itself conforms to Newtonian mechanics, the mechanics of such a wave pattern, considered as a particle located at its center, is much more complicated than that of the familiar mass point of particle dynamics. This complexity provides a bridge from the older concepts of particle behavior to the new.

The study of this model given below reveals no insuperable obstacles such as were encountered by the electromagnetic theory and the simpler ether model. The properties of the wave-patterns are qualitatively consistent with many of the concepts of modern physics, though in some cases not with the generality of application which is now assigned to them. Among these concepts are: the space-time of special relativity, relativistic mechanics, de Broglie waves, proportionality of energy and frequency, energy thresholds, and transfers of energy according to the quantum frequency formula. The ether model also leads to certain concepts not found in the present theories. It provides, for example, for a possible failure of the mass-energy balance such as has been observed in nuclear reactions. It also suggests the possibility of a new type of particle which, by virtue of its negative inertial mass, is capable of exerting a binding force between other particles.

These results make it more probable that classical mechanics may, after all, afford a sufficient "*foundation* for the physical description of all natural phenomena" even though the super-structure be very different from that contemplated by its originators. The present argument, however, is not that this particular description is necessary, but rather that it offers distinct advantages. On the philosophical side, there is the prospect of greater unification of the basic theory through a reduction in the number of independent assumptions. Matter and radiation appear as wave motions which satisfy the same equations. The apparent conflicts between current concepts appear to be reconcilable through a more exact determination of the conditions under which each applies. On the more practical side, the ether model provides a different approach and technique. It has the advantage inherent in all models that, once one is found which fits one set of conditions, a study of its properties under widely different conditions may bring out relations which it would be difficult to postulate solely on the basis of observations made under the second conditions. The suggested existence of particles having negative inertia, as discussed near the end of the paper, should it lead to anything of value, would be an example of such a relation. Also it makes available the added relationships which are characteristic of non-linear equations, without encountering those difficulties with respect to absolute motion which may arise when non-linearity is introduced arbitrarily. While the working out of the quantitative relations involved is a rather formidable undertaking, any effort in that direction may well throw new light on those problems which have not yielded to other methods.

The Gyrostatic Ether

As stated above the specific form of gyrostatic medium on which the present discussion is based is the ether model proposed by Kelvin. This is discussed in detail in a companion paper.[3] It is there shown that, for infinitesimal displacements, it is characterized by the wave equations:

$$\nabla \times \left(\frac{\bar{T}}{2}\right) = \rho_0 \frac{\partial \bar{q}}{\partial t} \tag{1}$$

$$\nabla \times \bar{q} = -\frac{1}{\eta_0} \frac{\partial}{\partial t} \left(\frac{\bar{T}}{2}\right), \tag{2}$$

where ρ_0 is the density, η_0 is a generalized stiffness determined by the constants of the medium, $\bar{q}$ is the vector velocity, and $\bar{T}$ is a vector torque per unit volume, which has its origin in the torque with which a gyrostat opposes an angular displacement of its axis. For a plane polarized plane wave, the quantity $\frac{\bar{T}}{2}$ can be interpreted as a surface tractive force per unit area, which a layer of the medium normal to the direction of propagation exerts on the layer just ahead. Its direction lies in the surface of separation, and is parallel to that of the velocity $\bar{q}$.

[3] R. V. L. Hartley, "The Reflection of Diverging Waves by a Gyrostatic Medium"—this issue of *The Bell System Technical Journal*.

These equations become identical with those of Maxwell for free space,

$$\nabla \times \bar{H} = \epsilon \frac{\partial \bar{E}}{\partial t},$$

$$\nabla \times E = -\mu \frac{\partial \bar{H}}{\partial t},$$

if we replace $\bar{q}$ by $\bar{E}$, $\frac{\bar{T}}{2}$ by $\bar{H}$, ρ_0 by ϵ and $\frac{1}{\eta}$ by μ. Then $\rho_0\bar{q}$ corresponds to $\bar{D}$ and -2φ to $\bar{B}$ where φ is the angular displacement of an element of the medium. Or the roles of the electric and magnetic quantities may be interchanged.

For present purposes, however, we are more interested in finite displacements. The relations which then apply are discussed in detail in the companion proper. It is there shown that changes of two kinds appear in (1) and (2), with corresponding changes in the transmission properties of the medium. The simple linear relations are to be replaced by non-linear ones, which cause distortion of a wave but no reflection. In addition, a qualitative difference appears in the nature of the elasticity, as was pointed out by Kelvin. The restoring torque is no longer proportional to the angular displacement alone. When the axis of a gyrostat is displaced it begins rotating toward the axis of the displacement, thereby decreasing the component of its spin which is normal to that axis. Thus the restoring torque for a constant angular displacement decreases with time. The restoring torque is therefore a function of the time as well as of the displacement. Because of this time dependence, a disturbance of finite amplitude generates waves which propagate both backward and forward.

For a plane progressive sine wave it is found that the reflected waves interfere destructively. However, if a central generator starts sending out a diverging sinusoidal disturbance, a part of the energy is reflected inward as a wave of the same frequency as the generator and another smaller part as waves the frequencies of which are odd multiples of that frequency. This reflection attenuates the outgoing wave. If the incoming wave is reflected rather than absorbed at the generator, it tends to set up a standing wave pattern. As time goes on, the impedance of the medium as seen from the generator becomes more reactive and less power is drawn from the generator. Due to the attenuation, the energy in spherical shells of a given thickness decreases with increasing radius, so that it and the power transmitted at the wave front approach zero as r approaches infinity. This falling off is somewhat similar to that suffered by a wave the frequency of which lies in the stop band of a filter, but with one important difference. There the attenuation is independent of the distance. But here, since the attenuation is a function of the magnitude of the disturbance and of the curvature of the wave-front, the attenuation constant approaches zero as r increases indefinitely.

Whether or not the total energy stored in the wave pattern will approach a finite or infinite value depends on how fast the attenuation decreases with distance, and a more complete solution is needed to give an exact answer. If it does approach infinity it will do so much more slowly than for a medium which does not reflect.

The disagreement between classical electromagnetics and mechanics, referred to above, may now be stated more explicitly. The former says that electromagnetic waves are represented exactly by Maxwell's equations, regardless of the magnitudes of the electromagnetic variables. When these waves are interpreted as existing in a mechanical ether, classical mechanics says that Maxwell's relationship is approached as a limit as the mganitudes approach zero. Waves of finite amplitude are to be represented by the more complicated relations.

The two systems differ in three important respects; their relation to uniform linear motion, the linearity of their equations and the nature of the elasticity involved. Because the classical electromagnetic equations are not invariant under a Newtonian transformation, the set of axes to which the equations refer are uniquely related to other sets which are moving uniformly with respect to them. In special relativity, this condition is avoided by modifying the classical concepts of space and time to conform to the fact that the equations are invariant under the Lorentz transformation. The Newtonian invariance of the ether equations, however, insures that a set of axes at rest with respect to the undisturbed ether is not unique. Hence in the modified model, in which *the motions which constitute matter* conform to the laws of the ether, a uniform linear velocity of the entire system cannot be detected. This is consistent with the accepted principle that absolute velocity is meaningless.

We are, however, still faced with the question of the detection of uniform motion of matter relative to the ether. This is discussed at length below, where it is shown that the properties of the ether lead directly to an auxiliary space-time, which applies very closely under the experimental conditions and accounts for the failure to detect the motion. This "experimental" space-time is formally identical with that of special relativity. Thus the modification of the space-time of classical electromagnetics which appears in special relativity might be said to bring it into closer formal agreement with the classical mechanics of ether wave patterns. At any rate the establishing of this theoretical connection between the space-time of special relativity and a classical mechanical model is a step toward unification.

On the matter of linearity, proposals have been made to add arbitrary non-linear terms to Maxwell's equations. While this also makes the electromagnetic equations more like those of the ether, an important difference still remains. An equation obtained in this way is not necessarily invariant under either a Newtonian or a Lorentz transformation. If, then, the axes with respect to which it is expressed are not to be unique, it must be shown that some transformation exists under which it is invariant. Not only is the form of the equation important here but also the interpretation of the dependent variables. For example, since the complete equations of the ether contain $q \cdot \nabla$, if the mechanical variables be replaced by the analogous electromagnetic ones, the equations will be Newtonian invariant only if E, which replaces $\bar{q}$, is interpreted as a velocity. It is evident, therefore, that the fact that we are dealing with a mechanical model is an important point in the argument. Also, unless the added terms make the effective constants depend on the time as well as the dependent variables, there will be no reflection of the energy in a finite disturbance and the medium will not have the energy trapping property which is essential to the present argument.

Stationary Wave Patterns

The first question to be considered is the possibility of setting up a sustained wave pattern suitable to represent a particle at rest with respect to the ether. The simplest procedure might seem to be to look for it as a solution of the approximate linear equations in the form of a pair of spherical waves propagating radially, one outward and one inward, so as to form together a standing wave pattern. However, certain difficulties are encountered. There is nothing in the free linear ether which can serve as boundary conditions to fix the position or size of the pattern. Even if these were determined, there would be nothing to fix the amplitude, and so the energy. Most patterns, particularly those which involve a single frequency, have one or more of the following features. Some of the variables become infinite at the center; the total energy is infinite, energy is propagated away radially.

These difficulties disappear, however, when we take account of the properties of the ether for disturbances of finite amplitude. Let us suppose that the energy which is to constitute the pattern is supplied by a central generator, the impedance of which is mainly reactive, so that reflected waves which reach it are reflected outward again. Once a standing wave pattern has been established as described above, let the force of the generator be reduced to zero without changing its impedance. The pattern will then persist except for a small and decreasing damping due to the outward radiation at its periphery. However, in the region near the center the displacements will be very large, and the incoming reflected waves will suffer reflections which increase with decreasing radius. These reflections will effectively take the place of the assumed reactive impedance of the generator, and so the latter may be discarded. The fact that the reflections take place from a somewhat diffuse inner boundary prevents the amplitude from building up to an infinite value at the center as it would with a linear medium.

However, the reflected wave includes components of triple and higher frequencies and, due to the non-linearity, other frequency components will be generated. If the entire pattern is to be stable, all of these must satisfy the boundary conditions. Their magnitudes relative to the fundamental, for a particular mode of oscillation, will depend on the amplitude and frequency of the fundamental, as well as on the constants of the medium. Hence the amplitude as well as the frequency of a stable pattern of a particular mode should be uniquely determined. Particles of different properties would then be expected to consist of patterns involving different modes of oscillation.

Returning to the lack of complete reflection at the outer boundary and the change it might be expected to make in the pattern with time, this might be an important factor for a single particle alone in the universe. Actually, however, a very large number of particles are present. If we consider a point at a considerable distance from any one particle, a point in a vacuum, the resultant of the disturbances produced there by all the patterns will be very large compared with that due to any one. But the effect on a particular pattern of its own loss by radiation will be determined by this small component, and so will be small compared with the effect exerted on it by the combined small fields of its neighbors. This combined field due to a large number of patterns, randomly placed, and moving at random, will constitute a randomly varying electromagnetic field in a vacuum, such as has recently been postulated for other reasons. If, now, the center of a pattern be placed at the point in question, this random field may occasionally take on so large a value as to disturb the equilibrium conditions of the pattern.

It may be argued that, in spite of the merging of a given pattern in that of the random group, the group as a whole will suffer a progressive loss of energy through incomplete reflection. Were this to occur the total loss of energy would not be evenly distributed among the particles. As discussed below the particles would exchange energy through the mechanism of the non-linearities, continually forming less stable group patterns of greater energy, which in turn suffer transitions to more stable patterns of lower energy. A small continuous decrease in total energy would manifest itself as an increase in the rate of transitions downward in energy compared to those upward.

Associated with a standing wave pattern such as that described above would be three regions. Near the center would be a relatively small core in which the non-linear effects predominate and linear theory is totally inapplicable. Farther out the departure from linearity is only moderate, and the variation of the constants with distance is slow enough that the reflections are small. It should be possible to treat wave propagation in this region by the methods developed for a string of variable density, which are sometimes cited as analogous with those employed in wave mechanics. The analogy is made closer by the fact that the variations in impedance which correspond to the varying density are determined by the energy density of the pattern itself. Still farther out the amplitudes become still smaller, the ether constants become very nearly but not quite uniform, and the pattern approaches very closely to that in a linear medium.

While the nature of the pattern is determined largely by the non-linear inner region, because of the small volume of this region most of the energy will be located in the nearly linear region. So we might expect some at least of the macroscopic properties of the pattern to differ very little from those deduced from a consideration of the corresponding pattern in a linear medium. We will therefore begin by examining such a pattern. For the linear case, when the axes are at rest with respect to the undisturbed ether, (1) and (2) lead to the wave equation for the vector displacement $\bar{s}$,

$$\frac{\partial^2 \bar{s}}{\partial t^2} = c^2 \nabla^2 \bar{s}. \tag{3}$$

As is well known, this is satisfied by any function of the form

$$\bar{s} = f(\omega t \pm k_x x \pm k_y y \pm k_z z),$$

where

$$\frac{\omega^2}{c^2} = k_x^2 + k_y^2 + k_z^2, \tag{4}$$

and the constants ω, k_x, k_y and k_z, are real or complex. Since an imaginary frequency is interpreted as an exponential change with time, it is not suitable for representing a permanent pattern, so ω will be taken to be real. Imaginary values of k are interpreted as exponential variations with distance. But, since $\bar{s}$ is always real, we may, by a four-dimensional Fourier analysis, represent f as the summation of components of the form

$$\bar{s} = \bar{A} \cos (\omega t \pm k_x x \pm k_y y \pm k_z z), \tag{5}$$

where $\bar{A}$ is a complex vector representing the amplitude and phase of the component, and k_x, k_y and k_z are real. Since each component must satisfy (3), the new constants must satisfy (4). Each such component constitutes a plane progressive wave traveling, with velocity c in a direction, the cosines of which are proportional to the wave numbers k_x, etc.

As a first step in building up a stationary pattern, in which there is no steady propagation of energy in any direction, we combine two progressive wave components (5) which are identical, except that their directions of phase propagation along, say, the z axis are opposite. The signs of the last terms are then opposite and the sum can be written

$$\bar{s} = 2\bar{A} \cos (\omega t \pm k_x x \pm k_y y) \cos k_z z.$$

Proceeding in the same way for x and y, we arrive at the standing wave pattern,

$$\bar{s} = 8\bar{A} \cos \omega t \cos k_x x \cos k_y y \cos k_z z. \tag{6}$$

Components of this sort, each with its own amplitude and phase, may be combined to build up possible stationary patterns. However, we shall not attempt here to build such patterns, but rather to deduce what information we can from a study of a single component.

Moving Wave Patterns

In order to represent approximately a particle in uniform linear motion, we are to look for a solution of (3) which represents a moving wave pattern. For this we make use of two functions which may readily be shown to be such solutions,

$$\bar{s} = g_+ \left(\beta(\omega + Vk_x)t - \beta \left(k_x + \frac{V\omega}{c^2} \right) x \pm k_y y \pm k_z z \right),$$

$$\bar{s} = g_- \left(\beta(\omega - Vk_x)t + \beta \left(k_x - \frac{V\omega}{c^2} \right) x \pm k_y y \pm k_z z \right),$$

where ω, k_x, k_y and k_z are real and satisfy (4), V is a real constant, and

$$\beta^2 = \frac{1}{1 - \frac{V^2}{c^2}}.$$

g_+ represents a plane progressive wave the propagation of which along the x axis is in the positive direction. g_- represents one of lower frequency, propagating in the negative x direction. Their wave numbers in the x direction differ in such a way that those in the y and z direction are the same for the two. In the plane wave case, where $k_y = k_z = 0$ and $\omega = ck_x$, they reduce to

$$\bar{s} = g_\pm \left(\beta \left(1 \pm \frac{V}{c} \right) \omega \left(t \mp \frac{x}{c} \right) \right).$$

The two waves then travel in the x direction with velocities c and $-c$, and their frequencies are in the ratio $\dfrac{c + V}{c - V}$.

In order to derive a quasi stationary pattern we replace the functions $g_+(\ \ )$ and $g_-(\ \ )$ by $\bar{B} \cos \alpha(\ \ )$ and combine components in a manner similar to that used in deriving (6). The result is

$$\bar{s} = 8\bar{B} \cos \alpha\beta\omega \left(t - \frac{V}{c^2} x \right) \cos \alpha\beta k_x(x - Vt) \cos \alpha k_y y \cos \alpha k_z z, \qquad (7)$$

where $\bar{B}$ is a complex vector, and α may be any real scalar function of V. When we compare this with (6) we find that the last three factors, which in (6) describe a fixed envelope, in (7) describe an envelope which moves in the x direction with velocity V. For the same values of k_x, k_y and k_z, the moving pattern has its dimensions in the x direction reduced relative to those in the y and z in the ratio $\frac{1}{\beta}$. The first factor in (6) describes a sinusoidal variation with time which is everywhere in the same phase. In (7) it describes one, the phase of which varies linearly with x. This factor also describes a wave which progresses in the x direction with a velocity $\frac{c^2}{V}$. The existence of such a wave as a factor in the expression for a moving wave pattern was commented on by Larmor.[4] Aside from the constant α in (7) it will be recognized as the Lorentz transform of (6), as it should be since the approximate equations of which it is a solution are invariant under this transformation.

We shall take (7) to represent one component of a moving wave pattern which represents a moving particle. If we transform this to axes moving with the pattern by a Newtonian transformation it becomes

$$\bar{s} = 8\bar{B} \cos \alpha \left(\frac{\omega}{\beta} t' - \frac{\beta\omega V}{c^2} x' \right) \cos \alpha\beta k_x x' \cos \alpha k_y y' \cos \alpha k_z z', \qquad (8)$$

in which the envelope is at rest. This may be thought of as a stationary wave in an ether which is moving relative to the axes with a velocity $-V$. It is a solution of the wave equation for such an ether, as obtained by transforming (3) to the moving axes, or

$$\frac{\partial^2 \bar{s}}{\partial t'^2} = c^2 \nabla'^2 \bar{s} + 2V \frac{\partial^2 \bar{s}}{\partial x' \partial t'} - V^2 \frac{\partial^2 \bar{s}}{\partial x'^2}.$$

The one dimensional form of this equation is identical with that given by Trimmer[5] for compressional waves in moving air, except that in one case $\bar{s}$ is solenoidal and in the other divergent.

So far we have found no reason to associate any particular moving pattern with the assumed stationary one, in the sense that the moving pat-

[4] Larmor, Ency. Brit. 11th Ed., 1910; 13th Ed., 1926, Vol. 22, p. 787.
[5] J. D. Trimmer, *Jour. Acous. Soc. Am.*, 9, p. 162, 1937.

tern describes the result of setting in motion the particle which is described by the stationary pattern. Without further knowledge or assumptions regarding the factors which control the form of the pattern, we can go no farther in this direction by theory alone. Rather than try to guess at these factors, it seems preferable to investigate what properties the wave patterns must have in order to conform to the known results of experiment.

Let us start with the Michelson-Morley experiment to which the earlier ether theory did not conform. The entire apparatus involved in the experiment is now to be considered as made up of particles each of which consists of a wave pattern in the ether. The apparatus as a whole may be regarded as a more complicated wave pattern. The interference pattern formed by the light beams may, if we wish, be included in the over-all pattern. The results to be expected in the experiment do not depend on the oscillatory nature of the wave, nor on its amplitude or phase, but only on its spatial distribution, which is determined by the envelope factors. It is obvious from (8) that, for any uniform velocity $-V$ of the ether relative to the apparatus, the ratios of the dimensions of the envelope along the motion to those across it are reduced, relative to their values when V is zero, in the ratio $\frac{1}{\beta}$. That is to say the apparatus like the fringes undergo this change in relative dimensions. But, as is well known, this is exactly what is required in order that there shall be no apparent motion of the fringes. Hence any one of the stationary patterns in a moving ether, as represented by (8), is consistent with the experiment. This experiment therefore furnishes no basis for selecting any particular pattern.

More generally, in any experiment, the distances and time intervals which are available as standards of comparison are associated with the wave patterns and change with their motion. Thus we may, following the special theory of relativity, define an auxiliary space and time, the units of which are associated with the dimensions and cyclic interval of a particular periodic wave pattern. This pattern then plays the roles of the "practically rigid body" and the "clock" which determine space and time in relativity theory. An examination of (8) shows that the dimensions of the pattern, its frequency, and its phase change with the velocity of the ether relative to the pattern in just the way that the corresponding quantities associated with the rigid body and clock change with velocity in the relativity theory. But there these changes are known to be such that no experiment can detect the velocity involved. It follows, therefore, that no experiment in which the apparatus consists of wave patterns of small amplitude is capable of detecting the velocity V, in (8), which in this case is the velocity of the ether relative to the apparatus. Hence any of the above patterns are consistent with the failure of all experiments designed to detect motion relative to the ether. When account is taken of the non-linearity of the ether the result to be expected should differ from that just found for the linear case only by the small difference between the linear and non-linear patterns, which may easily be too small to measure. Thus the principal obstacle to the older ether theory is removed.

While the special theory of relativity is usually written in the form which corresponds to α being unity in (8), it has long been recognized that there is no theoretical basis for this particular value. The ether patterns are consistent with the more general formulation. In order to pin down the value of α for the ether patterns we resort to another experiment. Ives and Stillwell[6] found that a molecule which emits radiation of frequency ω when at rest emits a frequency $\frac{\omega}{\beta}$ when in motion. This moving frequency is taken relative to axes moving with the molecule, and so is to be compared with the frequency of oscillation $\frac{\alpha}{\beta}\omega$ in (8). This indicates that in order to represent a component of the pattern which results when the fixed pattern is set in motion, we are to put α equal to unity.

Another observed relation is that the energy of a moving particle is β times that of the same particle at rest. This information should be useful in checking any theory of the mechanism by which the non-linearity of the medium determines the energy of the pattern. All we shall do here is to point out one relation, the significance of which from the standpoint of mechanism will be discussed below. In (7), where the frequency is expressed relative to the same axes as the energy of the moving pattern, if we put α equal to unity, the frequency also varies as β. Hence if the pattern conforms to experiment with respect to its energy, the energy must be proportional to the frequency.

Obviously, if we define the mass of the particle-pattern as its energy over c^2, the particle will conform to relativistic mechanics. The mass of a particle as so defined, while dimensionally the same as that of the ether, is in other respects quite different. Since it is derived from the energy associated with a disturbance of the ether, it would be zero in the undisturbed ether, while the ether mass would be finite. The momentum of a particle would be determined by the flow of energy associated with it. Also within a particle, if the mode of oscillation were such that the wave propagated continuously around the axis in one direction, the resulting rotation of the energy would be interpreted as an angular momentum or spin. This concept of spin was suggested by Japolsky[7] in connection with cylindrical waves in a linear medium. There is, therefore, no *a priori* reason to expect that the motion

[6] H. E. Ives and C. R. Stillwell, *Jour. Opt. Soc. Am.*, 28, 215, 1938 and 31, 369, 1941.
[7] N. S. Japolsky, *Phil. Mag.* 20, 417, 1935.

of particles should conform to the laws of classical mechanics. As just noted, it should conform much more closely to those of relativistic mechanics. Also, to the extent that the flow of energy follows the laws of wave mechanics, as suggested below, the behavior of the particles will also conform to those laws. Similar considerations apply to the mass of radiation as derived from its energy.

Another experiment which helps to fix the required properties of the patterns is that of Davisson and Germer, in which it is shown that a particle moving with velocity V is diffracted as if it had a wave length λ such that

$$\lambda = \frac{h}{\beta m_0 V},$$

where h is Planck's constant and m_0 is the rest mass.

If, in (7) with α unity, we assume the energy frequency ratio to be equal to h, the wavelength associated with the first factor reduces to the value given by experiment. This does not mean that an ordinary physical wave of this length is present in the pattern. It does mean that, at any instant, the amplitude of the sinusoidal variation of displacement with distance, as given by the remaining factors, varies sinusoidally with the wave length λ, and is zero as points separated by $\frac{\lambda}{2}$. Hence, when the presence of equally spaced obstacles calls for zero values of displacement at equally spaced intervals, the distorted wave should be capable of forming a stable diffraction pattern when the translational velocity of the pattern is such that the interval between points of zero displacement has the value required by the spacing of the obstacles.

Thus the wave pattern will conform to this experiment provided, first, that it is characterized by a particular wave length, and second, that the factor of proportionality between its energy and frequency is equal to h. The first requirement implies that the wave pattern when at rest has practically all of its energy associated with components which are all of the same frequency, or else are confined to a narrow band near the characteristic frequency.

At this point let us pause for a short review and discussion. Briefly, we have replaced the "rigid body" of special relativity by an oscillatory motion of the ether, the envelope of which is analogous with the configuration of the rigid body. We have found that when in motion this envelope behaves as does the rigid body, and the time relations conform to those of a moving clock. These latter may also be interpreted as a multiplying factor which has the form of a plane wave of the DeBroglie type. In wave mechanics, this is treated as a wave of a single frequency and of a variable phase velocity greater than that of light. In the ether theory this wave is interpreted as one factor in the description of an interference pattern which results from the superposition of component progressive waves of different frequencies, each of which travels with velocity c. This difference in viewpoint leads to other differences.

One of these has to do with the possibility of describing accurately both the position and velocity of a particle, which is ruled out from the wave mechanics viewpoint. An ether wave pattern, however, may have its position accurately described by its envelope, while at the same time the pattern moves with a definite velocity. The particle velocity may here be regarded as a group velocity derived from two waves progressing in opposite directions, but does not depend on the presence of dispersion as does that for waves in the same direction. It is not to be concluded from this that the position and velocity can be *measured* with this accuracy, for we have still to deal with the disturbing effect of the measurement.

From the ether viewpoint, one of the limitations of wave mechanics is to be expected, its inability to calculate directly the position of a particle. The information regarding this position is contained in the expression for the envelope, while the wave factor depends only on its state of motion. A calculation based on a solution which involves the wave factor without the envelope would be expected to be indefinite regarding position. We should expect, however, that it would give information as to the probability of the presence of the particle in a given region, since this is derivable from its state of motion.

Returning to the comparison with experiment, while wave patterns based on the linear equations have shown close agreement so far, the next experiment upsets the applecart. It has been observed that the motion of one particle is modified by the presence of other particles in its neighborhood. So long as the assumed equations are linear, the law of superposition holds, and every solution is independent of every other one. So any wave pattern, when once set up, will continue in its state of rest or of uniform motion indefinitely, and will not be influenced by the presence of other patterns or of free progressive waves. But these together comprise all other matter and radiation. Hence, while we have provided for the property of inertia, there is nothing which tends to alter the state of motion of a body, that is, there are no forces. In this respect the present linear treatment is similar to the special theory of relativity. So, in order to represent the interactions between particles, account must be taken of those between patterns which result from the non-linearity and time dependence of the ether.

Reactions between Patterns

The general problem of the effect of one pattern on another is even more intricate than that of the stable state of a single pattern, which it includes, and its solution will not be attempted here. Some conclusions may, however, be drawn. Since the amount of reflected energy generated by an element of the medium depends on powers of the instantaneous disturbance higher than the first, the superposition of a second pattern will alter the standing wave pattern of the first, and vice versa. Also, as pointed out in the companion paper, the propagation of both the main and reflected waves also depends on higher powers of the instantaneous disturbance there. The resulting variations in the propagation will also affect the conditions for a stable pattern. Neither pattern, then, can satisfy its stability conditions independently of the other; but if the combined patterns are to be stable they must together satisfy a new set of conditions common to both. How much each is altered by such a union will depend on the degree of coupling between them, that is, on the amount of energy which must be regarded as mutual to the two.

The effect of this coupling will be very different, depending on whether the frequencies of the two patterns are the same or different. When they are different the non-linear terms give rise to frequencies related to the first two by the quatum formula. The transfer of energy to these frequencies may, under favorable conditions, set up a new mode of oscillation the stability conditions of which are better satisfied than those of the original frequencies. The new mode might be that of an excited atom. Or the frequency of one or both of the patterns may be changed to that corresponding to the particle in motion with a particular velocity. In either of these processes some of the energy may be released as radiation at one of the difference frequencies.

If, however, the frequencies of the two patterns are identical, no new frequencies will result from their superposition. If the combined pattern is to persist there must be a stable mode for the combination, the frequency of which is identical with that of the separate patterns. This is hardly to be expected. Also the oscillations of the second pattern, being of the same frequency as those of the first, would have a much greater disturbing effect on its conditions for stability. It would appear, then, that if it were possible to bring two patterns of identical frequency into superposition, they would mutually disintegrate. This does not mean that two particles of the same type cannot exist in the same neighborhood. If they have different velocities, for example, their frequencies will be different. The similarity of these considerations to Pauli's exclusion principle is obvious.

If the second pattern has much greater energy than the first, as it will if it represents a much heavier particle, its stability conditions may be little affected by the presence of the first. The behavior of the first, an electron, may then be discussed on the assumption that it exists in a medium, the properties of which vary with position in accordance with the fixed pattern of the second particle, the nucleus. Since the stability conditions for the electron pattern particle are most strongly influenced by the effective constants of the medium near its center, we would expect its energy and frequency to be controlled largely by that part of the nuclear pattern which is near its center. Let us assume that, through some external agency, the center of the electron pattern is transferred from one position of rest to another which is differently placed relative to the nucleus. Owing to the different effect of the nuclear pattern on the effective constants of the medium as viewed by the electron pattern, the stable energy of the latter would be different at the second position. This change in rest energy with position may be interpreted as a measure of the change in a field of static potential associated with the massive nucleus. The similarity between this relationship and that which exists between the electron and the nuclear potential in wave mechanics is obvious.

In speaking of a change in the effective constants of the medium, we refer to an average value taken over a number of cycles and wave lengths of the oscillations which make up the second pattern, or nucleus. Calculations based on this concept should not therefore be expected to give valid results when the time intervals involved in the averages are comparable to the period $\frac{h}{m_0c^2}$ of the second particle at rest, or the distances are comparable to the corresponding wave length $\frac{h}{m_0c}$ of the pattern. For a proton this period is 4.38×10^{-24} seconds and the wave length is 1.31×10^{-13} cms. If, then, an electron is to be subject to the kind of nuclear potential field just described, the linear dimensions of that part of it which is controlled by the potential field of the proton must be at least of the order of 10^{-13} cm. This is consistent with Gamow's[8] observation that "It seems, in fact, that a length of the order of magnitude of 10^{-13} centimeters plays a fundamental role in the problem of elementary particles, popping out wherever we try to estimate their physical dimensions."

The variations in the medium due to the nucleus might be treated in terms of their effect on the progressive wave components, the interference of which gives rise to the wave pattern of the electron. The component waves as so influenced should combine to form an interference pattern which represents the behavior of the electron in the field of the nucleus. It is also possible that a technique may be found for treating their effect on that factor of the electron wave which is similar to the DeBroglie wave. This should be more nearly like the techniques now used in wave mechanics.

If two particles are brought so close together that the central cores of their patterns overlap, the departure from linearity becomes so great that

[8] G. Gamow, Physics Today, 2, p. 17, Jan., 1949.

a procedure which may be successful at intermediate separations becomes inadequate. Relativistic mechanics breaks down and Lorentz invariance may lose its significance. This is in agreement with the experimental result that, in some nuclear reactions, the energy balance, as calculated from the relativistic relations, is not satisfied. Also the difficulty which has been encountered in calculating nuclear phenomena by the techniques of wave mechanics suggests that the extremely non-linear condition is approached for the separation of the particles within a nucleus. This viewpoint suggests that an understanding of the nucleus might make possible an experimental determination of velocity relative to the ether.

The reactions between wave patterns of appreciable amplitude may also be viewed from a somewhat different angle. We may think of the various wave patterns as being the analogs of the various modes of motion of, say, an elastic plate. For very small amplitudes they have negligible effect on one another. For larger amplitudes, where Hooke's law does not hold, the force may be represented as a power series of the displacement. The first power term represents the linear stiffness. If the frequencies of two modes which are in oscillation are ω_1 and ω_2, the higher power terms represent forces of frequencies $m\omega_1 \pm n\omega_2$ where m and n are integers or zero. These forces set all the modes into forced oscillation at the frequencies of the various forces, in amounts which depend on the impedance of the particular mode for the particular frequency. When the frequency of the force coincides with the resonant frequency of one of the natural modes, the forced oscillations may be large. Thus the variation in stiffness with displacement provides a coupling whereby energy may be transferred from one or more modes, that is wave patterns, to other modes. But in this transfer the energy always appears associated with a new frequency which is related to those of the modes from which it came in accorance with the familiar formula of quantum theory.

The theory of such energy transformations with change of frequency has been worked out in considerable detail for vacuum tube and other variable resistance modulators, and the results show little in common with the quantum theory beyond the relations connecting the frequencies. When, however, the variation is not in a resistance but in a stiffness, as occurs in the ether case, the situation is quite different. This problem has been explored both theoretically[9] and experimentally.[10] It is found that an oscillation of one frequency in one mode may provide the energy to support sustained oscillations of two other lower frequencies in two other dissipative modes. For this to occur the frequencies involved must be related through the quantum formula. Also the amplitude of the generating oscillation must exceed a

[9] R. V. L. Hartley, *Bell Sys. Tech. Jour.*, 15, 424, 1936.
[10] L. W. Hussey and L. R. Wrathall, *Bell Sys. Tech. Jour.*, 15, 441, 1936.

threshold value which depends on the frequencies, the impedance involved, and the constant of non-linearity. The transformed energy divides itself between the generated modes in the ratio of their frequencies. In a non-dissipative system, the frequencies of possible combinations of sustained oscillations are determined by the energy of the system. Here also they are connected by the quantum formula.

The particle wave pattern discussed above would approximate very closely to such a non-dissipative non-linear system. We should therefore expect its frequency to be related to its energy through the constants of the ether. In the more complex wave patterns associated with more than one particle, it is unlikely that the pattern representing, say, an electron could maintain its identity as part of some arbitrarily chosen pattern, the magnitudes of which are not commensurable with its own. This suggests that the stable states of the complex pattern would be confined to a sequence of discreet patterns which are related to one another through some property of the electron. These possible non-dissipative combinations of energy and frequency would represent the stable quantum states of the atom. The radiation process would then be similar to that referred to above in which energy from a source of higher frequency distributes itself between two lower frequencies in the ratio of the frequencies. The energy in the pattern of an excited atom would serve as the source. One of the two lower frequencies would be that of a pattern corresponding to a lower energy state to which the transition occurs. The other would be that of the radiating wave which carries off the energy lost in the transition.

A Suggested New Particle

We saw above that the observed variation of the energy of a particle with its velocity calls for a mechanism in which the energy varies directly as the frequency. The fact that a system, in which the stiffness varies with the displacement, is characterized by this relation suggests that the energy of a particle pattern depends mainly on variations in the stiffness of the ether. However, the non-linearities of the ether equations cannot all be interpretated as variable stiffnesses. The non-linearity which appears in (1) when the displacements are finite is equivalent to a variable inertia. It is in order, therefore, to inquire into the properties of a pattern in which the energy is determined by this kind of non-linearity. The variable inductance of an iron-core coil constitutes such a variable inertia. Theoretical and experimental studies of circuits involving these coils have shown that they behave very much as do systems having variable stiffness, with one important exception. The energy distributes itself in the inverse ratio of the frequencies.

If, then, we assume that the energy of a moving pattern is determined by a mechanism which conforms to this relation, it follows from (7) that its energy will vary as $\frac{1}{\beta}$. Expanding in the usual manner we then have

$$W = m_0c^2 - \tfrac{1}{2}\, m_0V^2 + \cdots$$

This says that a particle represented by such a wave pattern would have a positive rest mass and a negative inertial mass. Its momentum is directed oppositely to its velocity, and energy must be taken from it to set it in motion and given to it to stop it. Such a particle, when bouncing back and forth between two rigid walls or rotating about two centers of force, would exert a force tending to draw them together, instead of the usual repulsion. It is interesting to speculate that if, in an atomic nucleus, the positive charges which are passed back and forth between other nuclear particles were associated with particles of this type their motion would exert a binding force on the other particles.

Conclusion

It appears, then, that the ether model is capable of sustaining wave patterns the behavior of which is qualitatively in agreement with the results of experiment. In order to establish fully the sufficiency of classical mechanics for the physical description of natural phenomena, it will be necessary to work out the complicated quantitative relations whereby the constants of the ether may be deduced from experimental measurements. However, until a serious attempt to do this has failed for some reason other than sheer mathematical complexity, the insufficiency of classical mechanics can scarcely be argued.

In conclusion, I wish to acknowledge the contributions of those of my colleagues who, through discussions over the years, have helped in developing the concepts which have been put together in the above picture.

The Reflection of Diverging Waves by a Gyrostatic Medium

By R. V. L. HARTLEY

(*Manuscript Received Feb. 28, 1950*)

This paper furnishes the basis for a companion one, which discusses the possibility of describing material particles as localized oscillatory disturbances in a mechanical medium. If a medium is to support such disturbances it must reflect a part of the energy of a diverging spherical wave. It is here shown that this property is possessed by a medium, such as that proposed by Kelvin, in which the elastic forces are of gyrostatic origin. This is due to the fact that, for a small constant angular displacement of an element of this medium, the restoring torque, instead of being constant, decreases progressively with time.

Introduction

In a companion paper[1] it is pointed out that it may be possible to describe the behavior of material particles as that of moving patterns of wave motion, provided a medium can be found which is capable of sustaining a localized oscillatory disturbance. In most media this is not possible, for the energy of the disturbance would be propagated away in all directions. Something special in the way of a medium is therefore called for. It must be capable of trapping the wave energy released from a central source. Kelvin proposed a mechanical medium, the equations of which, for small disturbances, were identical with those of Maxwell for free space. The medium derived its elasticity from gyrostats. He recognized that, for finite disturbances, the restoring torque depends on the time as well as the angular displacement. It is the present purpose to show that this time dependence imparts to his medium exactly the energy trapping property required.

The Gyrostatic Ether

The concept of an ether with stiffness to rotation originated with MacCullagh[2] in 1839, and was further developed by Kelvin[3] in 1888. MacCullagh showed that certain optical phenomena associated with reflection could not be represented by the elastic solid ether of Fresnel, but required for their mechanical representation a medium in which the potential energy is a function of what is now called the curl of the displacement. Fitzgerald[4] remarked in 1880 that its equations are identical with those of the electromagnetic

[1] R. V. L. Hartley, Matter, a Mode of Motion—this issue of the *Bell System Technical Journal.*

[2] Collected Works of James MacCullagh, Longmans Green & Co., London, 1880, p. 145.

[3] Mathematical and Physical Papers of Sir William Thomson, Vol. III, Art. XCIX, p. 436, and Art. C, p. 466.

[4] Phil. Trans. 1880, quoted by Larmor, Ether and Matter, Cambridge Univ. Press, 1900, p. 78.

theory of optics developed by Maxwell. This conclusion is confirmed in later discussion by Gibbs,[5] Larmor,[4] and Heaviside.[6]

Kelvin, apparently unaware of MacCullagh's work, was led by similar considerations to the same result. He went farther and devised a physical model which consisted of a lattice, the points of which were connected by extensible, massless, rigid rods in such a manner that the structure as a whole was incompressible and non-rigid. Each of these rods supported a pair of oppositely rotating gyrostats. By a gyrostat he meant a spinning rotor mounted in a gimbal so that it is effectively supported at its center of mass and can have its spin axis rotated by a rotation of the mounting. The resultant angular momentum of the rotors was the same in all directions.

This model, considered as a continuous medium, exhibits a stiffness to absolute rotation, the nature of which can be described by comparing it with the elasticity of a solid. A solid is characterized by a rigidity n such that small displacements u, v, w are accompanied by a stress tensor, one component of which is

$$n\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right).$$

For the ether model the corresponding component is

$$n\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) = 2n\varphi$$

where φ is a small angular displacement of the element about the z axis. More generally a small vector rotation $\overline{\Delta\varphi}$ is accompanied by a vector restoring torque per unit volume,

$$\overline{\Delta T} = -4n\overline{\Delta\varphi}. \tag{1}$$

The quantity $4n$ therefore represents a stiffness to angular displacement of the element.

In the appendix it is shown that the lattice of gyrostats, treated as a continuous medium, exhibits this kind of elasticity. It is also shown that for infinitesimal displacements, the medium is described by the wave equations (8a and 6a).

$$\nabla \times \left(\frac{\overline{T}}{2}\right) = \rho_0 \frac{\partial \bar{q}}{\partial t}, \tag{2}$$

$$\nabla \times \bar{q} = -\frac{1}{\eta_0}\frac{\partial}{\partial t}\left(\frac{\overline{T}}{2}\right), \tag{3}$$

[5] Collected Works of J. Willard Gibbs, Longmans Green & Co., New York 1928, Vol. II, p. 232.
[6] Heaviside, Electromagnetic Theory, Ernest Benn, Ltd., London, 1893, Vol. I, p. 226.

where ρ_0 is the constant density, η_0 is a generalized stiffness of the undisturbed medium, given by (7a), $\bar{q}$ is the vector velocity, and $\overline{T}$ is the torque per unit volume. In a plane wave $\bar{q}$ is normal to the direction of propagation. $\frac{\overline{T}}{2}$ is a tractive force per unit area in the direction of $\bar{q}$, which acts on a surface normal to the direction of propagation.

If, however, the amplitude is finite the equations become much more complicated. For present purposes we need consider only waves for which there is no component of velocity or torque in the direction of propagation, and we need consider only plane polarized waves for which the direction of the velocity is the same at all times and places. Also, as will appear below, we are concerned with the equations which describe a wave of infinitesimal amplitude which is superposed on a finite disturbance. This description need cover only infinitesimal ranges of time and position. It can therefore be expressed in terms of wave equations in which the constants of the medium have local instantaneous values which depend on the finite disturbance.

Subject to these restrictions it is shown in the appendix that (2) is to be replaced by (23a)

$$\nabla \times \left(\frac{\overline{T}}{2}\right) = l_q \rho \frac{\partial q}{\partial t}, \tag{4}$$

where l_q is a unit vector in the fixed direction of the velocity, and ρ is an instantaneous local density, defined in terms of the finite disturbance by (20a). And, in place of (3), (22a)

$$\nabla \times \bar{q} = -l_\varphi \frac{1}{\rho c^2}\left(\frac{\partial}{\partial t}\left(\frac{\overline{T}}{2}\right) + 2\frac{\partial f}{\partial t}\right), \tag{5}$$

where l_φ is a unit vector in the direction of the axis of rotation, ρ is again an instantaneous local density, c is an instantaneous local velocity derived in the usual way from ρ and an instantaneous local stiffness η, while f is a function defined by the relation, (13a),

$$\overline{T} = -l_\varphi 4f(\varphi, t).$$

This function takes account of the fact that when the spin axis of the rotor is given a constant finite displacement, the restoring torque is not constant as in (1), but changes with time as the spin axis rotates toward the axis of displacement, and so reduces the component of the spin which is normal to the displacement axis and so is effective in producing stiffness. $-4\frac{\partial f}{\partial t}$ represents the rate of this change in torque for a fixed angular displacement. $-4\frac{\partial f}{\partial \varphi}$ is to be interpreted as the rate of change of torque with angular displacement, when the time consumed is infinitesimal, that is when the angular velocity is infinite. It is therefore an instantaneous local angular stiffness from which the instantaneous local generalized stiffness η is derived as in (19a).

To simplify these expressions, let the direction of propagation be x and that of q be y. Then

$$\nabla \times \bar{q} = i \frac{\partial}{\partial x} (jq) = k \frac{\partial q}{\partial x},$$

so l_φ is in the direction of z, and represents a clockwise rotation about z. (5) then becomes the scaler equation

$$\frac{\partial q}{\partial x} = -\frac{1}{\rho c^2}\left[\frac{\partial}{\partial t}\left(\frac{T}{2}\right) + 2\frac{\partial f}{\partial t}\right]. \tag{6}$$

T is also in the z direction, so

$$\nabla \times \left(\frac{T}{2}\right) = i \frac{\partial}{\partial x}\left(k \frac{T}{2}\right) = -j \frac{\partial}{\partial x}\left(\frac{T}{2}\right).$$

But $\bar{q}$ is in the y direction, so

$$\frac{\partial}{\partial x}\left(\frac{T}{2}\right) = -\rho \frac{\partial q}{\partial t}. \tag{7}$$

These, then, are the desired equations of motion, for the type of wave under consideration.

The Generation of Reflected Waves

In this section we shall show that when a finite wave is propagated in this medium each element of the medium becomes the source of auxiliary waves which propagate in both directions from the source.

To do this we shall make use of the argument by which Riemann[7] showed that this does not occur for sound waves in an ideal gas. This will first be restated in more modern language. We consider a plane wave propagating along the x axis. We picture the finite pressure p and the longitudinal velocity u at a point in the medium as having been built up by the successive superposition of waves of infinitesimal amplitude, each propagating relative to the medium in its condition at the time of its superposition. If the first increment is propagating in the positive direction,

$$du = \frac{dp}{\rho c},$$

[7] Lamb, Hydrodynamics, Sixth Edition, p. 481. Rayleigh, Theory of Sound, Second Edition, Vol. II. p. 38.

where the characteristic resistance is ρc. Here

$$c^2 = \frac{dp}{d\rho}.$$

He assumes adiabatic expansion, so that p and c are functions of ρ only. If a second incremental wave of pressure dp, also traveling in the positive direction, be added, its velocity increment, being relative to the medium, will add to that already present. Its value will be related to dp through a new characteristic resistance corresponding to the modified density resulting from the previous increment. Hence the velocity u resulting from a large number of such waves will be

$$u = \int_0^p \frac{dp}{\rho c} = w,$$

where w is the quantity represented by ω in Lamb's version. If, then, all of the wave propagation is in the positive direction

$$u = w.$$

Similarly, if an incremental wave is traveling in the negative direction,

$$du = \frac{-dp}{\rho c},$$

and the condition for all the propagation to be in that direction is

$$u = -w.$$

Obviously, then, if u has some other value than one of these it results from the addition of increments some of which propagate in each direction.

Riemann deduces from the aerodynamic equations that

$$\left(\frac{\partial}{\partial t} + (u + c)\frac{\partial}{\partial x}\right)(w + u) = 0, \tag{8}$$

$$\left(\frac{\partial}{\partial t} + (u - c)\frac{\partial}{\partial x}\right)(w - u) = 0, \tag{9}$$

That is, the value of $w + u$ is propagated in the positive direction with a velocity of $c + u$ and that of $w - u$, in the negative direction with a velocity $c - u$. If, over a finite range of x, a disturbance be set up such that neither of these quantities is zero, it must be made up of incremental waves in both directions. However, as $w + u$ propagates positively it will be accompanied at any instant by a value of $w - u$ which has been propagated from the other direction. But, since the value of this was initially finite over a limited distance only, when all of this finite range is passed, $w - u$ will be zero, u will be equal to w and all of the wave will be traveling positively. A similar argument applies at the negative side of the wave. Thus the initial disturbance breaks up into two parts which travel in opposite directions without reflection. More generally, these considerations hold for any medium in which the stress is a function of the strain only.

For the ether model, since we have assumed the displacements are normal to the direction of propagation, the velocity of wave propagation relative to the medium is the same as that relative to the axes.

If now, following Riemann, we let

$$dw = \frac{1}{\rho c}\, d\left(\frac{T}{2}\right), \tag{10}$$

so that now

$$w = \int \frac{1}{\rho c}\, d\left(\frac{T}{2}\right),$$

then from (7) and (6)

$$\frac{\partial q}{\partial t} = -c\,\frac{\partial w}{\partial x},$$

$$\frac{\partial w}{\partial t} = -c\,\frac{\partial q}{dx} - \frac{2}{\rho c}\,\frac{\partial f}{dt}.$$

Adding and subtracting gives

$$\left(\frac{\partial}{\partial t} + c\,\frac{\partial}{\partial x}\right)(w + q) = -\frac{2}{\rho c}\,\frac{\partial f}{\partial t},$$

$$\left(\frac{\partial}{\partial t} - c\,\frac{\partial}{\partial x}\right)(w - q) = -\frac{2}{\rho c}\,\frac{\partial f}{\partial t},$$

which are to be compared with (8) and (9). Hence when $\dfrac{\partial f}{\partial t}$ is not zero the values of $w + q$ and $w - q$ are not propagated without change.

To show that reflection occurs, consider a disturbance at a point x at time t, characterized by q and w. At x and $t + \Delta t$, $w + q$ will differ from the value it had at $x - c\Delta t$, t, or $w + q - \dfrac{\partial}{\partial x}(w + q)c\Delta t$, by $-\dfrac{2}{\rho c}\dfrac{\partial f}{\partial t}\Delta t$. The increment at x in time Δt is

$$\Delta w + \Delta q = -\frac{\partial}{\partial x}(w + q)c\Delta t - \frac{2}{\rho c}\frac{\partial f}{\partial t}\Delta t,$$

and

$$\Delta w - \Delta q = \frac{\partial}{\partial x}(w - q)c\Delta t - \frac{2}{\rho c}\frac{\partial f}{\partial t}\Delta t.$$

From which

$$\Delta w = -c\frac{\partial q}{\partial x}\Delta t - \frac{2}{\rho c}\frac{\partial f}{\partial t}\Delta t,$$

$$\Delta q = -c\frac{\partial w}{\partial x}\Delta t.$$

Hence the velocity is the same as when $\frac{\partial f}{\partial t}$ is zero but w is changed by $-\frac{2}{\rho c}\frac{\partial f}{\partial t}\Delta t$. But the only way in which w can change with q constant is by adding waves of equal amplitude propagating in opposite directions, so that their contributions to w are equal and those to q are equal and opposite. From (10) this involves an increment of $\frac{T}{2}$ of $-2\frac{\partial f}{\partial t}\Delta t$ or a time rate of change of $-2\frac{\partial f}{\partial t}$. This agrees with (6), from which it is evident that the presence of $\frac{\partial f}{\partial t}$ alters $\frac{\partial q}{\partial x}$ from what it would otherwise be by $-\frac{2}{\rho c^2}\frac{\partial f}{\partial t}$. But, since q is unchanged, the velocities at $x + \frac{\Delta x}{2}$ and $x - \frac{\Delta x}{2}$ are increased by $-\frac{1}{\rho c^2}\frac{\partial f}{\partial t}\Delta x$ and $\frac{1}{\rho c^2}\frac{\partial f}{\partial t}\Delta x$. The first is the velocity associated with an auxiliary wave which propagates in the positive direction of x, and the second that of one which propagates in the negative direction, that is a reflected wave. Hence the medium generates a reflected wave of $\frac{1}{\rho c^2}\frac{\partial f}{\partial t}$ per unit length in the direction of propagation.

The Reflection of a Progressive Diverging Wave

So far attention has been confined to a single point. If a continuous disturbance is being propagated, it is important to know how the waves reflected at different points combine, for it is conceivable that they may interfere destructively. From the standpoint of the application to be made of these results in a companion paper, the case of most interest is that in which energy is propagated outward from a central generator as a sinusoidal wave of finite amplitude, beginning at time zero. Near the center, the wave of displacement will include radial as well as tangential components. As the radius increases the radial components become relatively negligible. We shall confine our attention to this outer region, where, in the absence of reflection, the propagation differs from that of a plane wave only in that the amplitude varies inversely as the radius. We shall neglect the effect of any reflections on the outgoing wave, and calculate the resultant reflected wave at a radius r_1 as a function of the time and so of the radial distance r the wave front has traveled.

If the outgoing wave were of infinitesimal amplitude, its velocity q_0 could be represented by

$$q_0 = \frac{r_0}{r} Q_0 \sin (\omega t - kr), \tag{11}$$

for values of $r < ct$, and by zero for $r > ct$, where Q_0 is the amplitude at some reference radius r_0. The sine function is chosen to avoid the necessity of an infinite acceleration at the wave front, as would be required by a cosine function. When the amplitude is finite this wave suffers distortion due to the fact that k which is equal to $\frac{\omega}{c}$ varies slightly with the variations in the instantaneous value of c. However, these will be small and, since fluctuations in velocity alone do not cause reflection, we shall neglect them. The procedure is to make use of q_0 to calculate the reflected wave increment generated in a length $\Delta r'$ at a radius r', calculate the amplitude and phase of this at a fixed point $r_1 < r'$, and at r_1 integrate the waves received there for values of r' from r_1 to the farthest point from which reflected waves can reach r_1 at the time t under consideration.

To find the reflected wave generated in a length $\Delta r'$ at r', we have from above that its velocity

$$\Delta q' = \frac{1}{\rho c^2} \frac{\partial f}{\partial t} \Delta r'.$$

From (21a), (19a) and (17a)

$$\frac{1}{\rho c^2} = \frac{F_1'}{\eta_0 \left(1 - a \left[\int \varphi \, dt\right]^2\right)},$$

where η_0 and a are constants of the medium given by (7a) and (15a). From (18a)

$$\frac{\partial f}{\partial t} = -a\eta_0\varphi^2 \int \varphi \, dt,$$

so

$$\frac{dq'}{dr'} = -\frac{aF_1'\varphi^2 \int \varphi\, dt}{1 - a\left[\int \varphi\, dt\right]^2}$$

which reduces to

$$\frac{dq'}{dr'} = -a\varphi^2 \int \varphi\, dt,$$

if we neglect second powers of the variables compared with unity.

To the same accuracy, from (14a)

$$\varphi = \frac{1}{2}\int \frac{\partial q_0}{\partial r'}\, dt.$$

From (11)

$$\frac{\partial q_0}{\partial r'} = -\frac{r_0 Q_0}{r'}\left[k \cos(\omega t - kr') + \frac{1}{r'}\sin(\omega t - kr')\right].$$

Here k is 2π over the wavelength so, if as we have assumed r_1, and therefore also r', is large compared with the wavelength, we may neglect the second term. Then

$$\varphi = -\frac{r_0 Q_0}{2cr'}\sin(\omega t - kr'),$$

$$\int \varphi\, dt = \frac{r_0 Q_0}{2c\omega r'}\cos(\omega t - kr'),$$

$$\frac{dq'}{dr'} = -\frac{a}{8\omega}\,\text{[illegible]}^3 \sin^2(\omega t - kr')\cos(\omega t - kr'),$$

$$= -\frac{a}{8\omega}\left(\frac{r_0 Q_0}{cr'}\right)^3[\cos(\omega t - kr') + \cos 3(\omega t - kr')].$$

This, when multiplied by $\Delta r'$, gives the value at r' of the wave, generated in the interval $\Delta r'$, which propagates in the negative direction of r. This is made up of components of frequency ω and 3ω. We are primarily interested, from the stand-point of reflection, in that of frequency ω, so we shall confine our attention to this component, with the understanding that the other can be treated in exactly the same fashion. As the fundamental component propagates inward to r_1 it increases in amplitude in the ratio $\frac{r'}{r_1}$ and suffers a phase lag of $k(r' - r_1)$. If we call the resultant of all the reflected waves at r_1, q_1', then the contribution to q_1' of the wave generated at r' is

$$\Delta q_1' = -\frac{a}{8\omega}\left(\frac{r_0 Q_0}{c}\right)^3 \frac{1}{r_1 r'^2} \cos(\omega t + kr_1 - 2kr')\Delta r'.$$

This is to be integrated from r_1 to the farthest point from which a reflected wave has reached r_1 at the instant t under consideration. This point is at $\frac{1}{2}(r_1 + ct)$. So

$$q_1' = -\frac{a}{8r_1\omega}\left(\frac{r_0 Q_0}{c}\right)^3 \int_{r_1}^{\frac{1}{2}(r_1+ct)} \frac{1}{r'^2} \cos(\omega t + kr_1 - 2kr')\, dr'.$$

Here the integrand is a function of r' and t and the upper limit of integration is also a function of t. We therefore make use of the relation[8]

$$\frac{d}{d\alpha}\int_a^b f(x,\alpha)\,dx = \int_a^b \left(\frac{\partial}{\partial\alpha} f(x,\alpha)\right) dx + f(b,\alpha)\frac{db}{d\alpha} - f(a,\alpha)\frac{da}{d\alpha}.$$

Putting t for α, r' for x we have

$$\frac{dq_1'}{dt} = \frac{a}{8r_1}\left(\frac{r_0 Q_0}{c}\right)^3 \left[\int_{r_1}^{\frac{1}{2}(r_1+ct)} \frac{1}{r'^2} \sin(\omega t + kr_1 - 2kr') - \frac{2c}{\omega}\frac{1}{(r_1 + ct)^2}\right]$$

which, upon integration becomes,

$$\frac{dq_1'}{dt} = \frac{a}{8r_1}\left(\frac{r_0 Q_0}{c}\right)^3 \Bigg(\frac{1}{r_1} si(\omega t + kr_1) - 2k[Si(\omega t - kr_1) - Si(2kr_1)]$$
$$\cdot \sin(\omega t + kr_1) - [Ci(\omega t + kr_1) - Ci(2kr_1)]$$
$$\cdot \cos(\omega t + kr_1) - \frac{2c}{\omega(r_1 + ct)^2}\Bigg).$$

Since q_1' is zero when t is $\frac{r_1}{c}$, its value at t will be found by integrating from $\frac{r_1}{c}$ to t, so

$$q_1' = \frac{a}{8r_1^2\omega}\left(\frac{r_0 Q_0}{c}\right)^3 \Bigg(-\cos(\omega t - kr_1) + \frac{2r_1}{r_1 + ct} - 2kr_1$$
$$\cdot \Bigg[\omega \int_{r_1/c}^{t} Si(\omega t + kr_1)\sin(\omega t + kr_1)\, dt + Si(2kr_1)$$
$$\cdot [\cos(\omega t + kr_1) - \cos 2kr_1] - \omega \int_{r_1/c}^{t} Ci(\omega t + kr_1)\cos(\omega t + kr_1)\, dt$$
$$+ Ci(2kr_1)[\sin(\omega t + kr_1) - \sin 2kr_1]\Bigg]\Bigg).$$

which reduces to

[8] Byerly, Integral Calculus, second edition p. 99.

$$q_1' = -\frac{a}{8r_1^2\omega}\left(\frac{r_0 Q_0}{c}\right)^3 \Bigg(\cos(\omega t - kr_1) - \frac{2r_1}{r_1 + ct} + 2kr_1$$
$$\cdot \, [-[Si(\omega t + kr_1) - Si(2kr_1)] \cos(\omega t + kr_1)$$
$$- [Ci(\omega t + kr_1) - Ci(2kr_1)] \sin(\omega t + kr_1)$$
$$+ Si(2\omega t + 2kr_1) - Si(4kr_1)]\Bigg).$$

The first term represents the value at r_1 of an outwardly moving wave in phase quadrature with the main wave. The second is a transient, the value of which is equal and opposite to that of the first term at the instant that the main wave passes r_1. The first two terms in the inner bracket are waves which propagate inward and so are to be regarded as reflections of the main wave. The last two terms represent a velocity which is zero when the main wave passes r_1, and subsequently oscillates about and approaches $\frac{\pi}{2} - \mathrm{Si}(4kr_1)$. Physically it appears to result from the particular form chosen for the main wave, which starts abruptly as a sine wave. The time integral of the impressed force, and so the applied momentum, has a component in one direction. Presumably if the main wave built up gradually these terms would be absent.

Returning to the reflected waves, their amplitudes are zero when the main wave passes r_1, after which they become finite. $Si(x)$ and $Ci(x)$ oscillate about and approach $\frac{\pi}{2}$ and zero respectively as x approaches infinity. Hence, as t increases indefinitely, the amplitudes of the reflected waves approach $\frac{\pi}{2} - Si(2kr_1)$ and $Ci(2kr_1)$. For the assumed large values of $2kr_1$ these quantities are small compared with unity. When multiplied by $2kr_1$ their variation is very slow. Hence the amplitudes vary roughly as $\frac{1}{r_1^2}$, and approach zero as the main wave at r_1 approaches an ideal plane one.

However, the significant fact is not that the reflected waves are small but that they are of finite magnitude. Because of this the main wave will not behave exactly as we assumed above, but will decrease slightly more rapidly with increasing radius. This should increase the reflection slightly, for the existence of the reflected wave is dependent on the decrease in amplitude with distance when the radius of curvature is finite.

To describe exactly what happens when the generator begins sending out waves from a central point would be hopelessly complicated, but we may form a general picture. In the early stages where the curvature is considerable, the reflected waves would be quite large and the main wave would be correspondingly attenuated. The arrival of the reflected waves at the generator adds a reactive component to the impedance of the medium, as seen from the generator, which reduces the power delivered to the medium. Meanwhile energy is being stored as standing waves in the medium and the rate of flow of energy in the wavefront is decreasing. The energy in successive shells of equal radial thickness decreases with increasing r, instead of being uniform as it would be in the absence of reflection. In the limit it approaches zero, but as the rate of decrease depends on the curvature, the rate of approach also approaches zero. As the rate at which energy is stored and that at which it is carried outward at the wavefront both approach zero, the resistance which the medium offers to the generator approaches zero, and its impedance approaches a pure reactance.

The total energy stored in the medium depends on how the over-all attenuation of the main wave is related to its amplitude. If there were no attenuation, the impedance would remain a pure resistance, the energy in successive shells would all be the same, and the total energy would increase linearly with r, and so with the time, and approach infinity. If the attenuation were independent of r, the total energy would approach a finite value. The present case is intermediate between these, the attenuation being finite but approaching zero with increasing r. If we assume it to vary as some power of the amplitude of the velocity, then W. R. Bennett has shown that if this power is less than the first the total energy approaches a finite value. If it is equal to the first, the energy approaches infinity as $\log r$, and if it is greater than this, the power approaches infinity more rapidly. Until more is known as to the actual variation of amplitude with distance, nothing definite can be said about the limit of the total energy.

APPENDIX: EQUATIONS OF THE KELVIN ETHER

We are concerned with the wave properties of the model for wavelengths long enough compared with the lattice constant so that it may be regarded as a continuous medium. Its density is equal to the average mass of the gyrostats per unit volume. Its elastic properties are to be derived from the resultant of the responses of the individual gyrostats.

We shall therefore begin by considering the behavior of a single element, which is shown schematically in Fig. 1. Here the outer ring of the gimbal, which is rigidly connected with the lattice, lies in the $x\,y$ plane. The axis about which the inner ring rotates is in the x direction, and the spin axis C of the rotor is in the z direction. We wish to examine the effect of a small angular displacement φ of the lattice, that is, of the outer ring. If it is about x or z, it will, because of the frictionless bearings, make no change in the rotor. If it is about y it will produce an equal displacement of the spin axis

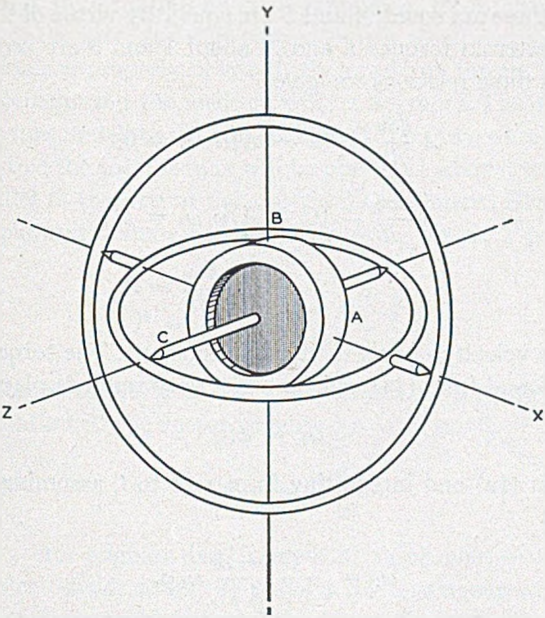


Fig. 1—Diagram of a gyrostat, showing its axes of rotation.

C about y. To study its effect we make use of Euler's equations for a rotating rigid body.[9]

$$A \frac{d\omega_1}{dt} - (B - C)\omega_2 \omega_3 = L,$$

$$B \frac{d\omega_2}{dt} - (C - A)\omega_3 \omega_1 = M,$$

$$C \frac{d\omega_3}{dt} - (A - B)\omega_1 \omega_2 = N,$$

where ω_1, ω_2 and ω_3 are the angular velocities about three principal axes of inertia, fixed in the rotor, the moments of inertia about which are A, B and C, and L, M, and N are the accompanying torques about the three axes. They are also at any instant the values of the torques about that set of axes, fixed in space, which, at the instant, coincide with the axes 1, 2, 3, which are fixed relative to the body. We let the 3 axis coincide with the spin axis C. We choose as the 1 and 2 axes, lines in the rotor which, at the instant, are in the x and y directions respectively. Since the moments of

[9] Jeans, Theoretical Mechanics, Ginn and Co., p. 308.

inertia about these are equal, A and B are equal. By virtue of the frictionless bearings the external torques L and N about 1 and 3 are zero.

Introducing these relations we have

$$A \frac{d\omega_1}{dt} + (C - A)\omega_2 \omega_3 = 0, \tag{1a}$$

$$A \frac{d\omega_2}{dt} - (C - A)\omega_1 \omega_3 = M, \tag{2a}$$

$$C \frac{d\omega_3}{dt} = 0. \tag{3a}$$

From (3a) the velocity of spin ω_3 remains constant. The torque M about y is then to be found from (1a) and (2a). For very small displacements,

$$\omega_2 = \dot{\varphi}.$$

Putting this in (1a) and integrating from zero to t, assuming φ to be zero at $t = 0$, gives

$$\omega_1 = -\frac{C - A}{A} \omega_3 \varphi.$$

(2a) then becomes

$$A\ddot{\varphi} + \frac{(C - A)^2}{A} \omega_3^2 \varphi = M.$$

This represents an angular inertia A and stiffness $\frac{(C - A)^2 \omega_3^2}{A}$. The system will therefore resonate at a frequency $\frac{(C - A)\omega_3}{A}$. If the frequencies involved in the variation of φ are small compared with this, the inertia torque will be negligible, and the system will behave as a stiffness. If the displacements about A associated with ω_1 are very small the restoring torque M will act substantially about the y axis. That is, the lattice will encounter a stiffness to rotation.

Since the large number of gyrostats in an element of the model are oriented in all directions, an angular displacement of the lattice about y will generally not be about the B axis for each gyrostat. If it makes an angle α with this axis, then only the component $\varphi \cos \alpha$ of the angular displacement will be transmitted to the rotor. The resulting torque will then be $S \cos \alpha$, where

$$S = \frac{(C - A)^2 \omega_3^2}{A}.$$

It will be directed about B and so will not be parallel to the applied displacement. However, if a second gyrostat has the position which the first would have if it were rotated about y through π, its torque along y is the same as that of the first, and that normal to it is equal and opposite. Hence, if the gyrostats are properly oriented, the resultant torque will be parallel to the displacement and the medium will be isotropic. The y component of the opposing torque will be $S\varphi \cos^2 \alpha$. Thus if the B axes are uniformly distributed in space the total torque will be one third what it would be if they were all parallel to the axis of the applied displacement. Hence if there are N gyrostats per unit volume the vector restoring torque $\overline{T}$ per unit volume will be

$$\overline{T} = -\frac{N}{3} \frac{(C - A)^2 \omega_3^2}{A} \bar{\varphi}. \tag{4a}$$

The next step is to derive the wave equations for a medium having this stiffness to rotation. If the vector velocity $\bar{q}$ is very small,

$$\nabla \times \bar{q} = 2 \frac{\partial \bar{\varphi}}{\partial t}, \tag{5a}$$

where $\bar{\varphi}$ is a vector angular displacement of an element of the medium at the point under consideration. 2φ plays a role analagous with that of the dilatation in compressional waves. Then, from (4a) and (5a),

$$\nabla \times \bar{q} = -\frac{1}{\eta_0} \frac{\partial}{\partial t} \left(\frac{\overline{T}}{2} \right), \tag{6a}$$

where the generalized stiffness of the undisturbed medium,

$$\eta_0 = \frac{N}{12} \frac{(C - A)^2}{A} \omega_3^2. \tag{7a}$$

To get the companion equation, we interpret the torque exerted by an element in terms of the forces it exerts on the surfaces of neighboring elements. Let the x axis Fig. 2 be in the direction of the torque $T\Delta x^3$ which is exerted by the medium within the small cube. This very small torque can be resolved into the sum of two couples, one consisting of an upward force $F_y \Delta x^2$ on the right face and an equal downward force on the left one, and the other of a leftward force $F_z \Delta x^2$ on the upper surface and a rightward one on the lower one. But, if there is not to be a shearing stress, F_y and F_z must be equal, and each equal to $\frac{T}{2}$. Thus a torque per unit volume T is equivalent to a set of tangential surface forces per unit area of $\frac{T}{2}$ each.

Now consider the force exerted on an element by its neighbors, through the adjoining surfaces. To take the simplest case, let T in Fig. 2 be everywhere in the x direction and independent of z but varying with y. Then the forces exerted on the upper and lower surfaces are equal and opposite. That downward on the right face exceeds that upward on the left by $\frac{\partial}{\partial y}\left(\frac{T}{2}\,\Delta x^2\right)\Delta y$, so the force in the z direction is $-\frac{\partial}{\partial y}\left(\frac{T}{2}\right)\Delta x^3$. By extending the argument to three dimensions it is easily shown that the total

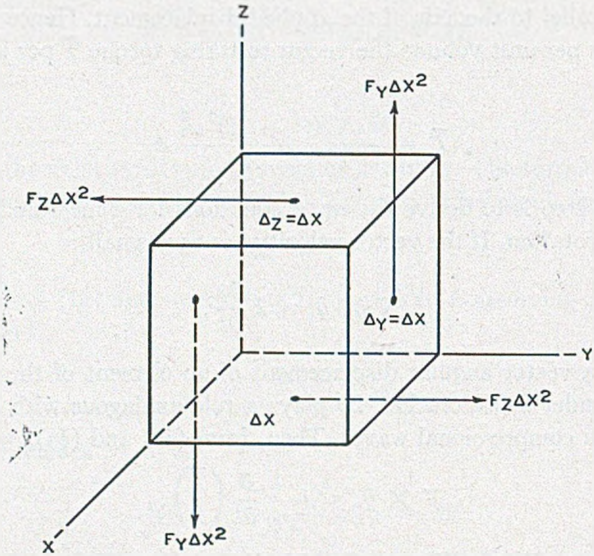


Fig. 2—Diagram showing the forces exerted by an element of the medium through its surfaces.

force is $\nabla \times \left(\frac{\overline{T}}{2}\right)\Delta x^3$. If ρ_0 is the density of the medium this force must equal $\rho_0 \Delta x^3 \frac{d\bar{q}}{dt}$, so

$$\nabla \times \left(\frac{\overline{T}}{2}\right) = \rho_0 \frac{d\bar{q}}{dt},$$

which, since $\bar{q}$ is small, reduces to

$$\nabla \times \left(\frac{\overline{T}}{2}\right) = \rho_0 \frac{\partial\bar{q}}{\partial t}. \tag{8a}$$

From this and (6a) the velocity of propagation is $(\eta_0/\rho_0)^{1/2}$ and the characteristic resistance is $(\rho_0\eta_0)^{1/2}$. In a plane wave the displacement is normal to the direction of propagation. The stress is a tractive force per unit area $\frac{T}{2}$ acting in a surface normal to the direction of propagation. It is in the direction of the velocity and in phase with it.

However, we are also interested in the case where the amplitudes are not negligible. We shall confine our attention to those cases where, as in plane or spherical waves at a distance from the source, the velocity is normal to the direction of propagation and the variations in the plane of the wave front are negligible. (5a) then becomes much more complicated. $\nabla \times \bar{q}$ is, however, still a function of $\frac{\partial \bar{\varphi}}{\partial t}$, say $2F_1\left(\frac{\partial \bar{\varphi}}{\partial t}\right)$. Then, for small variations of $\frac{\partial \bar{\varphi}}{\partial t}$ in the neighborhood of a particular value, we may write

$$\nabla \times \bar{q} = 2F_1'\left(\frac{\partial \bar{\varphi}}{\partial t}\right)_1 \frac{\partial \bar{\varphi}}{\partial t} \tag{9a}$$

where $F_1'\left(\frac{\partial \bar{\varphi}}{\partial t}\right)$ is a function of the particular value of $\frac{\partial \bar{\varphi}}{\partial t}$. This relation is to take the place of (5a). Similarly, if

$$\nabla \times \left(\frac{\bar{T}}{2}\right) = F_2\left(\frac{\partial \bar{q}}{\partial t}\right),$$

then, in place of (8a), we are to use, for small variations,

$$\nabla \times \left(\frac{\bar{T}}{2}\right) = F_2'\left(\frac{\partial \bar{q}}{\partial t}\right)\frac{\partial \bar{q}}{\partial t}. \tag{10a}$$

When we come to the transition from (5a) to (6a), however, the situation is somewhat different. To see how this comes about, we go back to the behavior of the single gyrostat of Fig. 1. It was assumed above that the B axis coincided with the y axis. However, when the displacement of the rotor about A is finite, this is no longer exactly true. The situation is then as shown in Fig. 3. A rotation φ of the lattice about y displaces A in the $x\,z$ plane by φ. The accompanying rotation of the rotor about A causes B to make an angle θ with y, which is independent of φ. Then

$$\omega_2 = \frac{d\varphi}{dt}\cos\theta.$$

From (1a)

$$\omega_1 = -\frac{C - A}{A}\,\omega_3 \int \frac{d\varphi}{dt}\cos\theta\, dt.$$

Also

$$\begin{aligned}\theta &= \int \omega_1\, dt,\\ &= -\frac{C - A}{A}\,\omega_3 \iint \frac{d\varphi}{dt}\cos\theta\, dt\, dt,\end{aligned} \tag{11a}$$

which determines θ as a function of φ and t. From (2a), neglecting the first term as above,

$$M = S \int \frac{d\varphi}{dt} \cos \theta \, dt,$$

and the restoring torque about y, or

$$T_y = -S \cos \theta \int \frac{d\varphi}{dt} \cos \theta \, dt. \tag{12a}$$

This, together with (11a), determines T_y as a function of φ and t, instead of φ alone as it is for infinitesimal displacements.

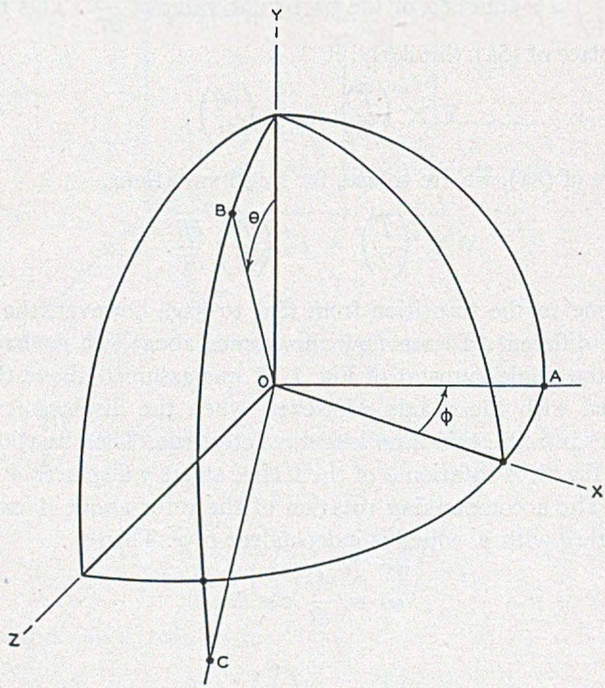


Fig. 3—Diagram showing the displacement of the axes of a gyrostat.

We assumed here that, in the rest position of the rotor, its B axis coincides with that of the applied displacement φ. When this is not the case, the relations are more complicated, but they should be qualitatively the same. Hence, for an element of the medium, the torque per unit volume should be a function of φ and t similar to T_y, which reduces to $-4\eta_0 \varphi$ for very small displacements. Since the restoring torque is in the direction of φ we may write

$$\bar{T} = -l_\varphi 4f(\varphi, t) \tag{13a}$$

where l_φ is a unit vector in the direction of the axis of rotation.

The derivation of the wave equation is much simpler if we consider only the case of present interest where the direction of the rotation is everywhere the same so that l_φ is constant. Then (9a) can be written as

$$\nabla \times \bar{q} = l_\varphi 2F_1'\left(\frac{\partial \varphi}{\partial t}\right)\frac{\partial \varphi}{\partial t}, \tag{14a}$$

and (13a) as

$$T = -4f(\varphi, t).$$

We wish now to replace $\dfrac{\partial \varphi}{\partial t}$ by $\dfrac{\partial}{\partial t}\left(\dfrac{T}{2}\right)$. These partial derivatives refer to a constant position so we are interested in the total time derivatives of T as given by (12a). To get the desired relation we need to express T explicitly in terms of φ and t, that is, we must evaluate φ. Since the variables are small, we neglect their products of higher order than the third. Then

$$\cos\theta = 1 - \frac{1}{2}a\left[\int \varphi\, dt\right]^2,$$

where

$$a = \left(\frac{C - A}{A}\omega_3\right)^2. \tag{15a}$$

Putting

$$T = -4\eta_0 \cos\theta \int \frac{d\varphi}{dt}\cos\theta\, dt,$$

in accordance with (12a) and substituting for $\cos\theta$ gives

$$T = -4\eta_0\left[\varphi - a\varphi\left[\int \varphi\, dt\right]^2 + a\int \varphi^2\left(\int \varphi\, dt\right)dt\right].$$

Then

$$\frac{dT}{dt} = -4\eta_0\left[\left(1 - a\left[\int \varphi\, dt\right]^2\right)\frac{d\varphi}{dt} - a\varphi^2\int \varphi\, dt\right].$$

When φ is constant the first term is zero, so the second term can be interpreted as the partial derivative of T with respect to t. Physically this describes the change in torque for a fixed displacement which results from the fact that, as the axis of the rotor rotates toward that of the applied torque, the component of the spin which is normal to the axis of displacement progressively diminishes. To interpret the first term, we let $\frac{d\varphi}{dt}$ increase indefinitely. The second term then becomes negligible, and when we divide through by $\frac{d\varphi}{dt}$, the left side becomes $\frac{dT}{d\phi}$. But the time increment which accompanies a finite increment of φ is now infinitesimal, and so this may be called the partial with respect to φ, with t constant.

We have then

$$\frac{dT}{dt} = -4\left(\frac{\partial f}{\partial\varphi}\frac{d\varphi}{dt} + \frac{\partial f}{\partial t}\right) \tag{16a}$$

where

$$\frac{\partial f}{\partial\varphi} = \eta_0\left(1 - a\left[\int\varphi\,dt\right]^2\right), \tag{17a}$$

$$\frac{\partial f}{\partial t} = -a\eta_0\varphi^2\int\varphi\,dt. \tag{18a}$$

Substituting for $\frac{\partial\varphi}{\partial t}$ from (16a) in (14a),

$$\nabla \times \bar{q} = -l_\varphi \frac{F_1'}{\frac{\partial f}{\partial\varphi}}\left(\frac{\partial}{\partial t}\left(\frac{T}{2}\right) + 2\frac{\partial f}{\partial t}\right).$$

We may interpret $\frac{\partial f}{\partial\varphi}$ as an instantaneous stiffness to rotation and define an instantaneous local generalized stiffness by the relation

$$\eta = \frac{\frac{\partial f}{\partial\varphi}}{F_1'}. \tag{19a}$$

Similarly from (10a) we may define an instantaneous density by the relation

$$\rho = F_2'. \tag{20a}$$

Then we may speak of an instantaneous velocity c given by

$$c^2 = \frac{\eta}{\rho}, \tag{21a}$$

and an instantaneous characteristic resistance ρc. Then

$$\nabla \times \bar{q} = -l_\varphi \frac{1}{\rho c^2}\left(\frac{\partial}{\partial t}\left(\frac{T}{2}\right) + 2\,\frac{\partial f}{\partial t}\right). \tag{22a}$$

(10a) becomes

$$\nabla \times \left(\frac{\overline{T}}{2}\right) = l_q \rho \frac{\partial q}{\partial t}, \tag{23a}$$

where l_q is a unit vector in the fixed direction of the velocity. These are the equations of motion which apply to a very small disturbance superposed on a finite disturbance.

Traveling-Wave Tubes

By J. R. PIERCE

[THIRD INSTALLMENT]

CHAPTER VII

EQUATIONS FOR TRAVELING-WAVE TUBE

Synopsis of Chapter

IN CHAPTER VI we have expressed the properties of a circuit in terms of its normal modes of propagation rather than its physical dimensions. In this chapter we shall use this representation in justifying the circuit equation of Chapter II and in adding to it a term to take into account the local fields produced by a-c space charge. Then, a combined circuit and ballistical equation will be obtained, which will be used in the following chapters in deducing various properties of traveling-wave tubes.

In doing this, the first thing to observe is that when the propagation constant Γ of the impressed current is near the propagation constant Γ_1 of a particular active mode, the excitation of that mode is great and the excitation varies rapidly as Γ is changed, while, for passive modes or for active modes for which Γ is not near to the propagation constant Γ_n, the excitation varies more slowly as Γ is changed. It will be assumed that Γ is nearly equal to the propagation constant Γ_1 of one active mode, is not near to the propagation constant of any other mode and varies over a small fractional range only. Then the sum of terms due to all other modes will be regarded as a constant over the range of Γ considered. It will also be assumed that the phase velocities corresponding to Γ and Γ_1 are small compared with the speed of light. Thus, (6.47) and (6.47a) are replaced by (7.1), where the first term represents the excitation of the Γ_1 mode and the second term represents the excitation of passive and "non-synchronous" modes. In another sense, this second term gives the field produced by the electrons in the absence of a wave propagating on the circuit, or, the field due to the "space charge" of the bunched electron stream. Equation (7.1) is the equation for the distributed circuit of Fig. 7.1. This is like the circuit of Fig. 2.3 save for the addition of the capacitances C_1 between the transmission circuit and the electron beam. We see that, because of the presence of these capacitances, the charge of a bunched electron beam will produce a field in addition to the field of a wave traveling down the circuit. This circuit is intuitively so appealing that it was originally thought of by guess and justified later.

Equation (7.1), or rather its alternative form, (7.7), which gives the voltage in terms of the impressed charge density, can be combined with the ballistical equation (2.22), which gives the charge density in terms of the voltage, to give (7.9), which is an equation for the propagation constant. The attenuation, the difference between the electron velocity and the phase velocity of the wave on the circuit in the absence of electrons and the difference between the propagation constant and that for a wave traveling with the electron speed are specified by means of the gain parameter C and the parameters d, b and δ. It is then assumed that d, b and δ are around unity or smaller and that C is much smaller than unity. This makes it possible to neglect certain terms without serious error, and one obtains an equation (7.13) for δ.

In connection with (7.7) and Fig. 7.1, it is important to distinguish between the *circuit voltage* V_c, corresponding to the first term of (7.7), and the total voltage V acting on the electrons. These quantities are related by (7.14). The a-c velocity v and the convection current i are given within the approximation made ($C \ll 1$) by (7.15) and (7.16).

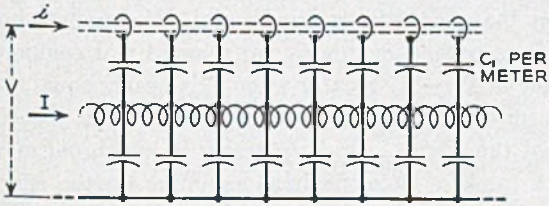


Fig. 7.1

7.1 Approximate Circuit Equation

From (6.47) we can write for a current $J = i$ and a summation over n modes

$$E_z = (1/2)(\Gamma^2 + \beta_0^2)i \sum_n \frac{(E^2/\beta^2 P)_n \Gamma_n^3}{(\Gamma_n^2 + \beta_0^2)(\Gamma_n^2 - \Gamma^2)} \tag{6.47a}$$

This has a number of poles at $\Gamma = \Gamma_n$. We shall be interested in cases in which Γ is very near to a particular one of these, which we shall call Γ_1. Thus the term in the expansion involving Γ_1 will change rapidly with small variations in Γ. Moreover, even if $(E^2/\beta^2 P)_1$ and Γ_1 have very small real components, $\Gamma_1^2 - \Gamma^2$ can be almost or completely real for values of Γ which have only small real components. Thus, one term of the expansion, that involving Γ_1, can go through a wide range of phase angles and magnitudes for very small fractional variations in Γ, fractional variations, as it turns out, which are of the order of C over the range of interest.

The other modes are either passive modes, for which even in a lossy circuit $(E^2/\beta^2 P)_n$ is almost purely imaginary, and Γ_n almost purely real, or they are active modes which are considerably out of synchronism with the electron velocity. Unless one of these other active modes has a propagation constant Γ_n such that $|(\Gamma_1 - \Gamma_2)/\Gamma_1|$ is so small as to be of the order of C, the terms forming the summation will not vary very rapidly over the range of variation of Γ which is of interest.

We will thus write the circuit equation in the approximate form

$$E = \left[\frac{\Gamma^2\Gamma_1(E^2/\beta^2P)}{2(\Gamma_1^2 - \Gamma^2)} - \frac{j\Gamma^2}{\omega C_1}\right] i \tag{7.1}$$

Here there has been a simplification of notation. E is the z component of electric field, as in Chapter II, and is assumed to vary as $\exp(-\Gamma z)$. (E^2/β^2P) is taken to mean the value for the Γ_1 mode. It has been assumed that β_0^2 is small compared with $|\Gamma_1^2|$ and $|\Gamma^2|$, and β_0^2 has been neglected in comparison with these quantities.

Further, it has been pointed out that for slightly lossy circuits, (E^2/β^2P) will have only a small imaginary component, and we will assume as a valid approximation that (E^2/β^2P) is purely real. We cannot, however, safely assume that Γ_1 is purely imaginary, for a small real component of Γ_1 can affect the value of $\Gamma_1^2 - \Gamma^2$ greatly when Γ is nearly equal to Γ_1.

The first term on the right of (7.1) represents fields associated with the active mode of the circuit, which is nearly in synchronism with the electrons. We can think of these fields as summing up the effect of the electrons on the circuit over a long distance, propagated to the point under consideration.

The term $(-j\Gamma^2/\omega C_1)$ in (7.1) sums up the effect of all passive modes and of any active modes which are far out of synchronism with the electrons. It has been written in this form for a special purpose; the term will be regarded as constant over the range of Γ considered, and C_1 will be given a simple physical meaning.

This second term represents the field resulting from the local charge density, as opposed to that of the circuit wave which travels to the region from remote points. Let us rewrite (7.1) in terms of voltage and charge density

$$E = -\frac{\partial V}{\partial z} = \Gamma V \tag{7.2}$$

From the continuity equation

$$i = (j\omega/\Gamma)\rho \tag{2.18}$$

$$V = \left[\frac{j\omega\Gamma_1(E^2/\beta^2P)}{2(\Gamma_1^2 - \Gamma^2)} + \frac{1}{C_1}\right]\rho \tag{7.3}$$

We see that C_1 has the form of a capacitance per unit length. We can, for instance, redraw the transmission-line analogue of Fig. 2.3 as shown in Fig. 7.1. Here, the current I is still the line current; but the voltage V acting on the beam is the line voltage plus the drop across a capacitance of C_1 farads per meter.

Consider as an illustration the case of unattenuated waves for which

$$\Gamma_1 = j\beta_1 \tag{7.5}$$

$$\Gamma = j\beta \tag{7.6}$$

where β_1 and β are real. Then

$$V = \left[\frac{\omega\beta_1(E^2/\beta^2P)}{2(\beta_1^2 - \beta^2)} + \frac{1}{C_1}\right]\rho \tag{7.7}$$

In (7.7), the first term in the brackets represents the impedance presented to the beam by the "circuit"; that is, the ladder network of Figs. 2.3 and 7.1. The second term represents the additional impedance due to the capacitance C_1, which stands for the impedance of the nonsynchronous modes. We note that if $\beta < \beta_1$, that is, for a wave faster than the natural phase velocity of the circuit, the two terms on the right are of the same sign. This must mean that the "circuit" part of the impedance is capacitive. However, for $\beta > \beta_1$, that is, for a wave slower than the natural phase velocity, the first term is negative and the "circuit" part of the impedance is inductive. This is easily explained. For small values of β the wavelength of the impressed current is long, so that it flows into and out of the circuit at widely separated points. Between such points the long section of series inductance has a higher impedance than the shunt capacitance to ground; the capacitive effect predominates and the circuit impedance is capacitive. However, for large values of β the current flows into and out of the circuit at points close together. The short section of series inductance between such points provides a lower impedance path than does the shunt capacitance to ground; the inductive impedance predominates and the circuit impedance is inductive. Thus, for *fast* waves the circuit appears *capacitive* and for *slow* waves the circuit appears *inductive*.

Since we have justified the use of the methods of Chapter II within the limitations of certain assumptions, there is no reason why we should not proceed to use the same notation in the light of our fuller understanding. We can now, however, regard V not as a potential but merely as a convenient variable related to the field by (7.2).

From (2.18) and (7.3) we obtain

$$V = \left[\frac{\Gamma\Gamma_1(E^2/\beta^2P}{2(\Gamma_1^2 - \Gamma^2)} - \frac{j\Gamma}{\omega C_1}\right]i \tag{7.8}$$

We use this together with (2.22)

$$i = \frac{jI_0\beta_e\Gamma V}{2V_0(j\beta_e - \Gamma)^2} \tag{2.22}$$

We obtain the overall equation

$$1 = \frac{jI_0\beta_e\Gamma}{2V_0(j\beta_e - \Gamma)}\left[\frac{\Gamma\Gamma_1(E^2/\beta^2P)}{2(\Gamma_1 - \Gamma)} - \frac{j\Gamma}{\omega C_1}\right] \tag{7.9}$$

In terms of the gain parameter C, which was defined in Chapter II,

$$C^3 = (E^2/\beta^2P)(I_0/8V_0) \tag{2.43}$$

we can rewrite (7.8)

$$(j\beta_e - \Gamma)^2 = \frac{j2\beta_e\Gamma^2\Gamma_1C^3}{(\Gamma_1^2 - \Gamma^2)} + \frac{4\beta_e\Gamma^2C^3}{\omega C_1(E^2/\beta^2P)} \tag{7.10}$$

We will be interested in cases in which Γ and Γ_1 differ from β_e by a small amount only. Accordingly, we will write

$$-\Gamma = -j\beta_e + \beta_eC \tag{7.11}$$

$$-\Gamma_1 = -j\beta_e - j\beta_eCb - \beta_eCd \tag{7.12}$$

The propagation constant Γ describes propagation in the presence of electrons. A positive real value of δ means an increasing wave. A positive imaginary part means a wave traveling faster than the electrons.

The propagation constant Γ_1 refers to propagation in the circuit in the absence of electrons. A positive value of b means the electrons go faster than the undisturbed wave. A positive value d means that the wave is an attenuated wave which decreases as it travels.

If we use (7.11) and (7.12) in connection with (7.10) we obtain

$$\delta = \frac{[1 + C(2j\delta - C\delta^2)][1 + C(b - jd)]}{[-b + jd + j\delta + C(jbd - b^2/2 + d^2/2 + \delta^2/2)]} - \frac{4\beta_e[(1 + C(2j\delta - C\delta^2)]C}{\omega C_1(E^2/\beta^2P)} \tag{7.13}$$

We will now assume that $|\delta|$ is of the order of unity, that $|b|$ and $|d|$ range from zero to unity or a little larger, and that $C \ll 1$. We will then neglect the parentheses multiplied by C, obtaining

$$\delta = \frac{1}{(-b + jd + j\delta)} - 4QC \tag{7.14}$$

$$Q = \frac{\beta_e}{\omega C_1(E^2/\beta^2P)} \tag{7.15}$$

The quantity ωC_1 has the dimensions of admittance per unit length, β_e has the dimensions of (length)$^{-1}$ and $(E^2/\beta^2 P)$ has the dimensions of impedance. Thus, Q is a dimensionless parameter (the space-charge parameter) which may be thought of as relating to the impedance parameter $(E^2/\beta^2 P)$ associated with the synchronous mode the impedance $(\beta_e/\omega C_1)$, attributable to all modes but the synchronous mode.

At this point it is important to remember that there are not only two impedances, but two voltage components as well. Thus, in (7.8), the first term in the brackets times the current represents the "circuit voltage", which we may call V_c. The second term in the brackets represents the voltage due to space charge, the voltage across the capacitances C_1. The two terms in the brackets are in the same ratio as the two terms on the right of (7.14), which came from them. Thus, we can express the circuit component of voltage V_c in terms of the total voltage V acting on the beam either from (7.8) as

$$V_c = \left[1 - \frac{j2(\Gamma_1^2 - \Gamma^2)}{\omega C_1 \Gamma_1 (E^2/\beta^2 P)}\right]^{-1} V \tag{7.16}$$

or, alternatively, from (7.14) as

$$V_c = [1 - 4QC(-b + jd + j\delta)]^{-1} V \tag{7.17}$$

From Chapter II we have relations for the electron velocity (2.15) and electron convection current (2.22). If we make the same approximations which were made in obtaining (7.14), we have

$$(ju_0 C/\eta)v = \frac{V}{\delta} \tag{7.18}$$

$$(-2V_0 C^2/I)i = \frac{V}{\delta^2} \tag{7.19}$$

We should remember also that the variation of all quantities with z is as

$$e^{-j\beta_e z} e^{\beta_e C \delta z} \tag{7.20}$$

The relations (7.18)–(7.19) together with (2.36), which tells us that the characteristic impedance of the circuit changes little in the presence of electrons if C is small, sum up in terms of the more important parameters the linear operation of traveling-wave tubes in which C is small. The parameters are: the gain parameter C, relative electron velocity parameter b, circuit attenuation parameter d and space-charge parameter Q. In following chapters, the practical importance of these parameters in the operation of traveling-wave tubes will be discussed.

There are other effects not encompassed by these equations. The effect of transverse electron motions is small in most tubes because of the high focusing fields employed; it will be discussed in a later chapter. The differences between a field theory in which different fields act on different electrons and the theory leading to (7.14)–(7.20), which apply accurately only when all electrons at a given z-position are acted on by the same field, will also be discussed.

CHAPTER VIII

THE NATURE OF THE WAVES

Synopsis of Chapter

IN THIS CHAPTER we shall discuss the effect of the various parameters on the rate of increase and velocity of propagation of the three forward waves. Problems involving boundary conditions will be deferred to later chapters.

The three parameters in which we are interested are those of (7.13), that is, b, the velocity parameter, d, the attenuation parameter and QC, the space-charge parameter. The fraction by which the electron velocity is greater than the phase velocity for the circuit in the absence of electrons is bC. The circuit attenuation is 54.6 dC db/wavelength. Q is a factor depending on the circuit impedance and geometry and on the beam diameter. For a helically conducting sheet of radius a and a hollow beam of radius a_1, Q can be obtained from Fig. 8.12.

The three forward waves vary with distance as

$$e^{-j\beta_e(1-yC)z}e^{\beta_e xCz}$$

$$\beta_e = \frac{\omega}{u_0}$$

Thus, a positive value of y means a wave which travels faster than the electrons, and a positive value of x means an increasing wave. The gain in db per wavelength of the increasing waves is BC, and B is defined by (8.9).

Figure 8.1 shows x and y for the three forward waves for a lossless circuit ($d = 0$). The increasing wave is described by x_1, y_1. The gain is a maximum when the electron velocity is equal to the velocity of the undisturbed wave, or, when $b = 0$. For large positive values of b (electrons much faster than undisturbed wave), there is no increasing wave. However, there is an increasing wave for all negative values of b (all low velocities). For the increasing wave, y_1 is negative; thus, the increasing wave travels more slowly than the electrons, *even when the electrons travel more slowly than the circuit wave in the absence of electrons*. For the range of b for which there is an increasing wave, there is also an attenuated wave, described by $x_2 = -x_1$ and $y_2 = y_1$. There is also an unattenuated wave described by $y_3(x_3 = 0)$.

For very large positive and negative values of b, the velocity of two of the waves approaches the electron velocity (y approaches zero) and the velocity of the third wave approaches the velocity of the circuit wave in the absence of electrons (y approaches minus b). For large negative values of b, x_1, y_1 and x_2, y_2 become the "electron" waves and y_3 becomes the "circuit" wave. For large values of b, y_1 and y_3 become the "electron" waves and y_2 becomes the "circuit" wave. The "circuit" wave is essentially the wave in the absence of electrons, modified slightly by the presence of a non-synchronous electron stream. The "electron waves" represent the motion of "bunches" along the electron stream, slightly affected by the presence of the circuit.

Figures 8.2 and 8.3 indicate the effect of loss. Loss decreases the gain of the increasing wave, adds to the attenuation of the decreasing wave and adds attenuation to the wave which was unattenuated in the lossless case. For large positive and negative values of b, the attenuation of the circuit wave (given by x_3 for negative values of b and x_2 for positive values of b) approaches the attenuation in the absence of electrons.

Figure 8.4 shows B, the gain of the increasing wave in db per wavelength per unit C. Figure 8.5 shows, for $b = 0$, how B varies with d. The dashed line shows a common approximation: that the gain of the increasing wave is reduced by $\frac{1}{3}$ of the circuit loss. Figure 8.6 shows how, for $b = 0$, x_1, x_2 and x_3 vary with d. We see that, for large values of d, the wave described by x_2 has almost the same attenuation as the wave on the circuit in the absence of electrons.

Figures 8.7–8.9 show x, y for the three waves with no loss ($d = 0$) but with a-c space charge taken into account ($QC \neq 0$). The immediately striking feature is that there is now a minimum value of b below which there is no increasing wave.

We further note that, for large negative and positive values of b, y for the electron waves approaches $\pm 2\sqrt{QC}$. In these ranges of b the electron waves are dependent on the electron inertia and the field produced by a-c space charge, and have nothing to do with the active mode of the circuit.

As QC is made larger, the value of b for which the gain of the increasing wave is a maximum increases. Now, C is proportional to the cube root of current. Thus, as current is increased, the voltage for maximum gain of the increasing wave increases. An increase in optimum operating voltage with an increase in current is observed in some tubes, and this is at least partly explained by these curves.* There is also some decrease in the maximum value of x_1 and hence of B as QC is increased. This is shown more clearly in Fig. 8.10.

If x and B remained constant when the current is varied, then the gain per wavelength would rise as C, or, as the $\frac{1}{3}$ power of current. However,

* Other factors include a possible lowering of electron speed because of d-c space charge, and boundary condition effects.

we see from Fig. 8.10 that B falls as QC is increased. The gain per wavelength varies as BC and, because Q is constant for a given tube, it varies as BQC. In Fig. 8.11, BQC, which is proportional to the gain per wavelength of the increasing wave, is plotted vs QC, which is proportional to the $\frac{1}{3}$ power of current. For very small values of current (small values of QC), the gain per wavelength is proportional to the $\frac{1}{3}$ power of current. For larger values of QC, the gain per wavelength becomes proportional to the $\frac{1}{4}$ power of current.

It would be difficult to present curves covering the simultaneous effect of loss (d) and space charge (QC). As a sort of substitute, Figs. 8.13 and 8.14 show $\partial x_1/\partial d$ for $d = 0$ and b chosen to maximize x_1, and $\partial x_1/\partial(QC)$ for $QC = 0$ and $b = 0$. We see from 8.13 that, while for small values of QC the gain of the increasing wave is reduced by $\frac{1}{3}$ of the circuit loss, for large values of QC the gain of the increasing wave is reduced by $\frac{1}{2}$ of the circuit loss.

8.1 Effect of Varying the Electron Velocity

Consider equation (7.13) in case $d = 0$ (no attenuation) and $Q = 0$ (neglect of space-charge). We then have

$$\delta^2(\delta + jb) = -j \tag{8.1}$$

Here we will remember that

$$\beta_e = \omega/u_0 \tag{8.2}$$

$$-\Gamma_1 = -j\beta_e(1 + Cb) = -j\omega/v_1 \tag{8.3}$$

Here v_1 is the phase velocity of the wave in the absence of electrons, and u_0 is the electron speed. We see that

$$u_0 = (1 + Cb)v_1 \tag{8.4}$$

Thus, $(1 + Cb)$ is the ratio of the electron velocity to the velocity of the *undisturbed wave*, that is, the wave in the absence of electrons. Hence, b is a measure of velocity difference between electrons and undisturbed wave. For $b > 0$, the electrons go faster than the undisturbed wave; for $b < 0$ the electrons go slower than the undisturbed wave. For $b = 0$ the electrons have the same speed as the undisturbed wave.

If $b = 0$, (8.1) becomes

$$\delta^3 = -j \tag{8.5}$$

which we obtained in Chapter II.

In dealing with (8.1), let

$$\delta = x + jy$$

The meaning of this will be clear when we remember that, in the presence of electrons, quantities vary with z as (from (7.10))

$$\begin{aligned} &e^{-j\beta_e(1+jC\delta)z} \\ &= e^{-j\beta_e(1-Cy)z}\, e^{\beta_e Cxz} \end{aligned} \tag{8.6}$$

If v is the phase velocity in the presence of electrons, we have

$$\omega/v = (\omega/u_0)(1 - Cy) \tag{8.7}$$

If $Cy \ll 1$, very nearly

$$v = u_0(1 + Cy) \tag{8.8}$$

In other words, if $y > 0$, the wave travels faster than the electrons; if $y < 0$ the wave travels more slowly than the electrons.

From (8.6) we see that, if $x > 0$, the wave increases as it travels and if $x < 0$ the wave decreases as it travels. In Chapter II we expressed the gain of the increasing wave as

$$BCN \text{ db}$$

where N is the number of wavelengths. We see that

$$\begin{aligned} B &= 20(2\pi)(\log_{10}e)x \\ B &= 54.5x \end{aligned} \tag{8.9}$$

In terms of x and y, (8.1) becomes

$$(x^2 - y^2)(y + b) + 2x^2y + 1 = 0 \tag{8.10}$$

$$x(x^2 - 3y^2 - 2yb) = 0 \tag{8.11}$$

We see that (8.11) yields two kinds of roots: those corresponding to unattenuated waves, for which $x = 0$ and those for which

$$x^2 = 3y^2 + 2yb \tag{8.12}$$

If $x = 0$, from (8.10)

$$\begin{aligned} y^2(y + b) &= 1 \\ b &= -y + 1/y^2 \end{aligned} \tag{8.13}$$

If we assume values of y ranging from perhaps $+4$ to -4 we can find the corresponding values of b from (8.13), and plot out y vs b for these unattenuated waves.

For the other waves, we substitute (8.12) into (8.10) and obtain

$$2yb^2 + 8y^2b + 8y^3 + 1 = 0 \tag{8.14}$$

This equation is a quadratic in b, and, by assigning various values of y, we can solve for b. We can then obtain x from (8.12).

In this fashion we can construct curves of x and y vs b. Such curves are shown in Fig. 8.1.

We see that for

$$b < (3/2)(2)^{1/3}$$

there are two waves for which $x \neq 0$ and one unattenuated wave. The increasing and decreasing waves ($x \neq 0$) have equal and opposite values of x, and since for them $y < 1$, they travel more slowly than the electrons, *even when the electrons travel more slowly than the undisturbed wave.* It can be

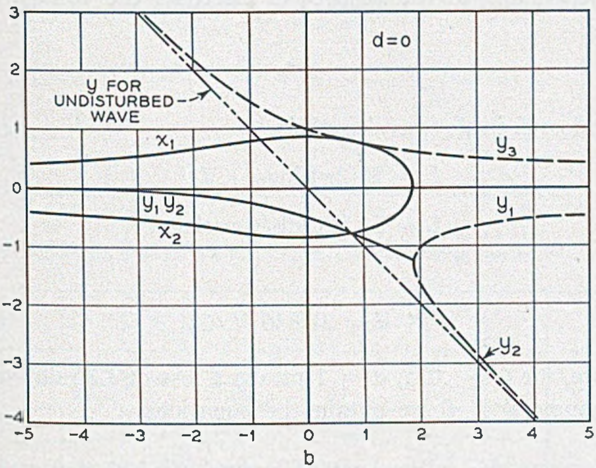


Fig. 8.1—The three waves vary with distance as exp $(-j\beta_e + j\beta_e Cy + \beta_e Cx)z$. Here the x's and y's for the three waves are shown vs the velocity parameter b for no attenuation ($d = 0$) and no space charge ($QC = 0$).

shown that the electrons must travel faster than the increasing wave in order to give energy to it.

For $b > (3/2)(2)^{1/3}$, there are 3 unattenuated waves: two travel faster than the electrons and one more slowly.

For large positive or negative values of b, two waves have nearly the electron speed ($|y|$ small) and one wave travels with the speed of the undisturbed wave. We measure velocity with respect to electron velocity. Thus, if we assigned a parameter y to describe the velocity of the undisturbed wave relative to the electron velocity, it would vary as the 45° line in Fig. 8.1.

The data expressed in Fig. 8.1 give the variation of gain per wavelength of the undisturbed wave with electron velocity, and are also useful in fitting boundary conditions; for this we need to know the three x's and the three y's.

In a tube in which the total gain is large, a change in b of $\pm$ 1 about $b = 0$ can make a change of several db in gain. Such a change means a difference between phase velocity of the undisturbed wave, v_1, and electron velocity u_0 by a fraction approximately $\pm C$. Hence, the allowable difference between phase velocity v_1 of the undisturbed wave, which is a function of frequency, and electron velocity, which is not, is of the order of C.

8.2 Effect of Attenuation

If we say that $d \neq 0$ but has some small positive value, we mean that the circuit is lossy, and in the absence of electrons the voltage decays with distance as

$$e^{-\beta_e C d}$$

Hence, the loss L in db/wavelength is

$$\begin{aligned} L &= 20(2\pi)(\log_{10} e)Cd \\ L &= 54.5Cd \text{ db/wavelength} \end{aligned} \tag{8.15}$$

or

$$d = .01836\ (L/C) \tag{8.16}$$

For instance, for $C = .025$, $d = 1$ means a loss of 1.36 db/wavelength.

If we assume $d \neq 0$ we obtain the equations

$$(x^2 - y^2)(y + b) + 2xy(x + d) + 1 = 0 \tag{8.17}$$

$$(x^2 - y^2)(x + d) - 2xy(y + b) = 0 \tag{8.18}$$

The equations have been solved numerically for $d = .5$ and $d = 1$, and the curves which were obtained are shown in Figs. 8.2 and 8.3. We see that for a circuit with attenuation there is an increasing wave for all values of b (electron velocity). The velocity parameters y_1 and y_2 are now distinct for all values of b.

We see that the maximum value of x_1 decreases as loss is increased. This can be brought out more clearly by showing x_1 vs b on an expanded scale. It is perhaps more convenient to plot B, the db gain per wavelength per unit C, vs b, and this has been done for various values of d in Fig. 8.4.

We see that for small values of d the maximum value of x_1 occurs very near to $b = 0$. If we let $b = 0$ in (8.17) and (8.18) we obtain

$$y(x^2 - y^2) + 2xy(x + d) + 1 = 0 \tag{8.19}$$

$$(x^2 - y^2)(x + d) - 2xy^2 = 0 \tag{8.20}$$

We can rewrite (8.20) in the form

$$y = \pm x \left(\frac{1 + d/x}{3 + d/x} \right)^{1/2} \tag{8.21}$$

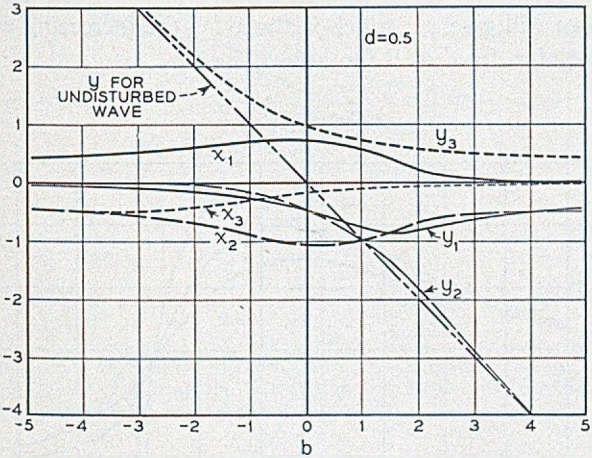


Fig. 8.2—The x's and y's for a circuit with attenuation ($d = .5$).

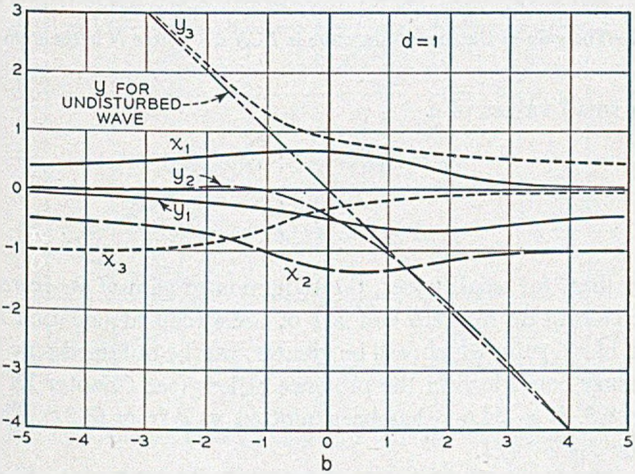


Fig. 8.3—The x's and y's for a circuit with attenuation ($d = 1$).

If we substitute this into (8.19) we can solve for x in terms of the parameter d/x

$$x = \mp \left[\frac{\left(\frac{3 + d/x}{1 + d/x} \right)^{1/2}}{2\left(\frac{1}{3 + d/x} + 1 + d/x \right)} \right]^{1/3} \tag{8.22}$$

Here we take both upper signs or both lower signs in (8.21) and (8.22).

If we assume $d/x \ll 1$ and expand, keeping no powers of d/x higher than the first, we obtain

$$x = \mp (\sqrt{3}/2)(1 - (1/3(d/x)) \tag{8.23}$$

The plus sign will give x_1, which is the x for the increasing wave. Let x_{10} be the value of x_1 for $d = 0$ (no loss).

$$x_{10} = \sqrt{3}/2 \tag{8.24}$$

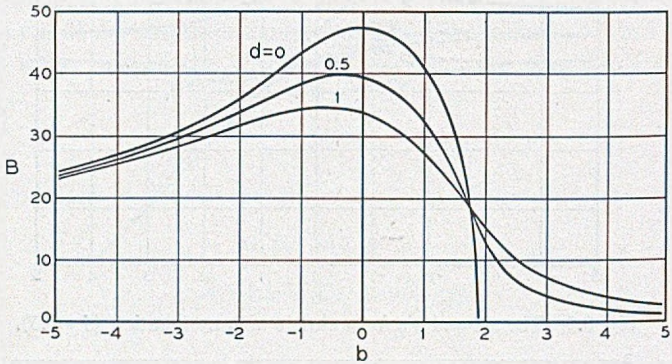


Fig. 8.4—The gain of the increasing wave is BCN db, where N is the number of wavelengths.

Then for small values of d

$$\begin{aligned} x_1 &= x_{10}(1 - (1/3)(d/x_{10})) \\ x_1 &= x_{10} - 1/3d \end{aligned} \tag{8.25}$$

This says that, for small losses, the reduction of gain *of the increasing wave* from the gain in db for zero loss is $\frac{1}{3}$ of the circuit attenuation in db. The reduction of *net gain*, which will be greater, can be obtained only by matching boundary conditions in the presence of loss (see Chapter IX).

In Fig. 8.5, $B = 54.6\ x_1$ has been plotted vs d from (8.22). The straight line is for $x_{10} = d/3$.

In Fig. 8.6, $-x_1$, x_2 and x_3 have been plotted vs d for a large range in d. As the circuit is made very lossy, the waves which for no loss are unattenuated and increasing turn into a pair of waves with equal and opposite small attenuations. These waves will be essentially disturbances in the electron stream, or space-charge waves. The original decreasing wave turns into a wave which has the attenuation of the circuit, and is accompanied by small disturbances in the electron stream.

8.3 Space-Charge Effects

Suppose that we let d, the attenuation parameter, be zero, but consider cases in which the space-charge parameter QC is not zero. We then obtain

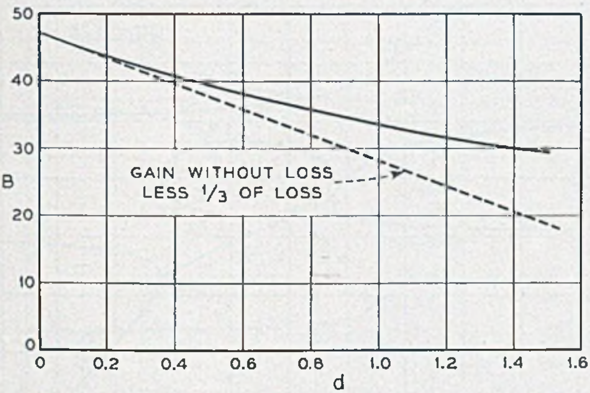


Fig. 8.5—For $b = 0$, that is, for electrons with a velocity equal to the circuit phase velocity, the gain factor B falls as the attenuation parameter d is increased. For small values of d, the gain is reduced by $\frac{1}{3}$ of the circuit loss.

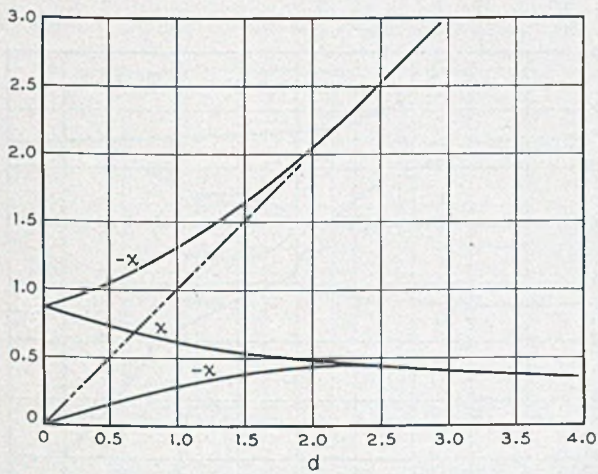


Fig. 8.6—How the three x's vary for $b = 0$ and for large losses.

the equations

$$(x^2 - y^2)(b + y) + 2x^2y + 4QC(b + y) + 1 = 0 \tag{8.26}$$

$$x[(x^2 - y^2) - 2y(y + b) + QC] = 0 \tag{8.27}$$

Solutions of this have been found by numerical methods for $QC = .25$, .5 and 1; these are shown in Figs. 8.7–8.9.

We see at once that the electron velocity for maximum gain shifts markedly as QC is increased. Hence, the region around $b = 0$ is not in this case worthy of a separate investigation.

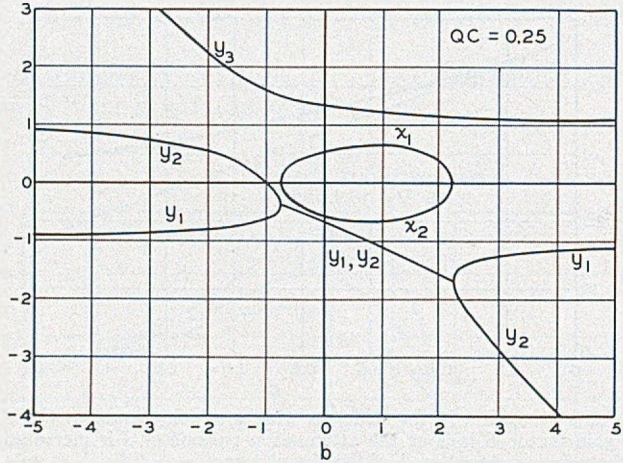


Fig. 8.7—The x's and y's for the three waves with zero loss ($d = 0$) but with space charge ($QC = .25$).

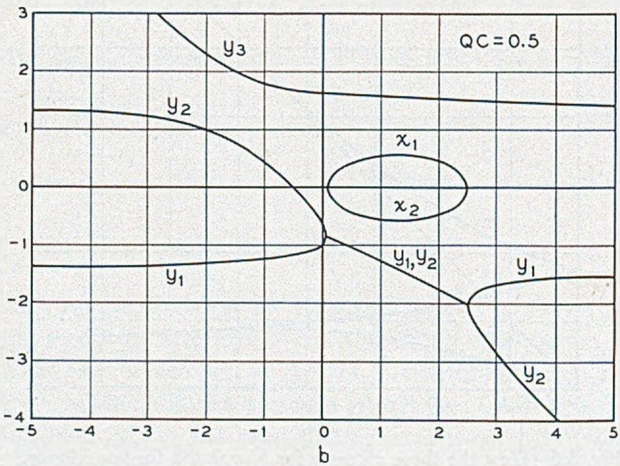


Fig. 8.8—The x's and y's with greater space charge ($QC = .5$).

It is interesting to plot the maximum value of x_1 vs. the parameter QC. This has, in effect, been done in Fig. 8.10, which shows B, the gain in db per wavelength per unit C, vs. QC.

We can obtain a curve proportional to db per wavelength by plotting BQC vs. QC. (Q is independent of current.) This has been done in Fig. 8.11. For $QC < 0.025$, the gain in db per wavelength varies linearly with

QC. Chu and Rydbeck found that under certain conditions gain varies approximately as the $\frac{1}{4}$ power of the current. This would mean a slope of $\frac{3}{4}$ on Fig. 8.11. A $\frac{3}{4}$ power dashed line is shown in Fig. 8.11; it fits the upper part of the curve approximately.

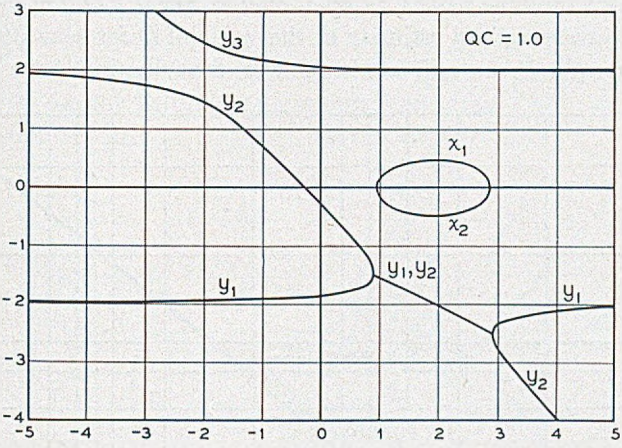


Fig. 8.9—The x's and y's with still greater space charge ($QC = 1$).

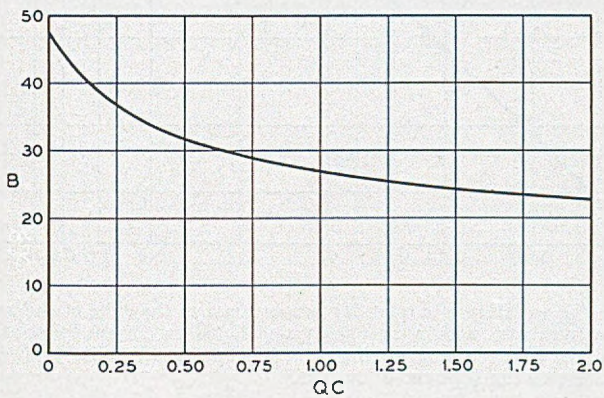


Fig. 8.10—How the gain factor B decreases as QC is increased, for the value of b which gives a maximum value of x_1.

If we examine Figs. 8.7–8.9 we find that for large and small values of b there are, as in other cases, a circuit wave, for which y is nearly equal to $-b$, and two space-charge waves. For these, however, y does not approach zero.

Let us consider equation (7.13). If b is large, the first term on the right becomes small, and we have approximately

$$\delta = \pm j2\sqrt{QC} \tag{8.28}$$

These waves correspond to the space-charge waves of Hahn and Ramo, and are quite independent of the circuit impedance, which appears in (8.28) merely as an arbitrary parameter defining the units in which δ is measured. Equation (8.28) also describes the disturbance we would get if we shorted out the circuit by some means, as by adding excessive loss.

Practically, we need an estimate of the value of Q for some typical circuit. In Appendix IV an estimate is made on the following basis: The helix

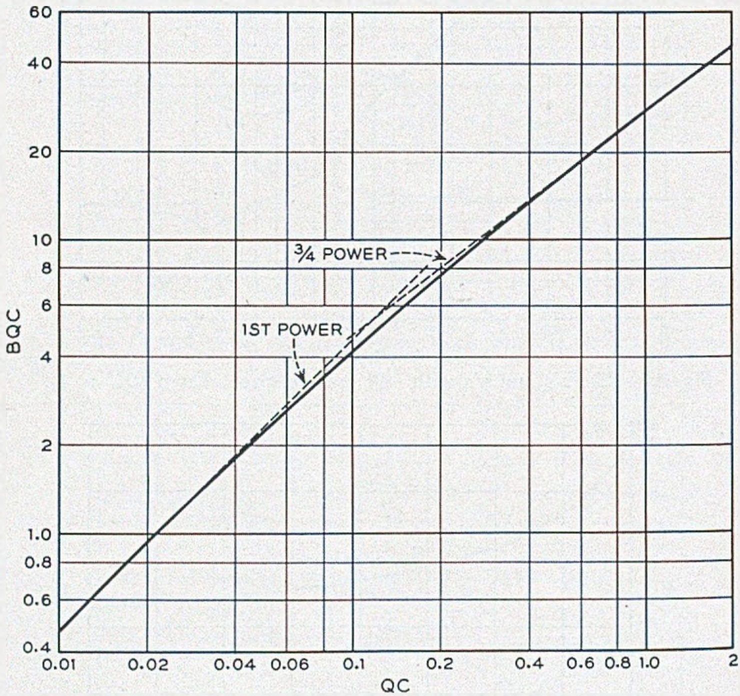


Fig. 8.11—The variation of a quantity proportional to the cube of the gain of the increasing wave (ordinate) with a quantity proportional to current (abscissa). For very small currents, the gain of the increasing wave is proportional to the $\frac{1}{3}$ power of current, for large currents to the $\frac{1}{4}$ power of current.

of radius a is replaced by a conducting cylinder of the same radius, a thin cylinder of convection current of radius a_1 and current of $i \exp(-j\beta z)$ is assumed, and the field is calculated and identified with the second term on the right of (7.1). R. C. Fletcher has obtained a more accurate value of Q by a rigorous method. His work is reproduced in Appendix VI, and in Fig. 1 of that appendix, Fletcher's value of Q is compared with the approximate value of Appendix IV.

In Fig. 8.12, the value $Q(\beta/\gamma)^2$ of Appendix IV is plotted vs. γa for $a_1/a = .9, .8, .7$. For $a_1/a = 1$, $Q = 0$. In a typical 4,000 mc traveling-wave tube, $\gamma a = 2.8$ and C is about .025. Thus, if we take the effective beam radius as .5 times the helix radius, $Q = 5.6$ and $QC = .14$.

We note from (7.14) that Q is the ratio of a capacitive impedance to $(E^2/\beta^2 P)$. In obtaining the curves of Fig. 8.12, the value of $(E^2/\beta^2 P)$ for a helically conducting sheet was assumed. This is given by (3.8) and (3.9). If $(E^2/\beta^2 P)$ is different for the circuit actually used, and it is somewhat different, even for an actual helix, Q from Fig. 8.12 should be multiplied by $(E^2/\beta^2 P)$ for the helically conducting sheet, from (3.8) and (3.9), and divided by the value of $(E^2/\beta^2 P)$ for the circuit used.

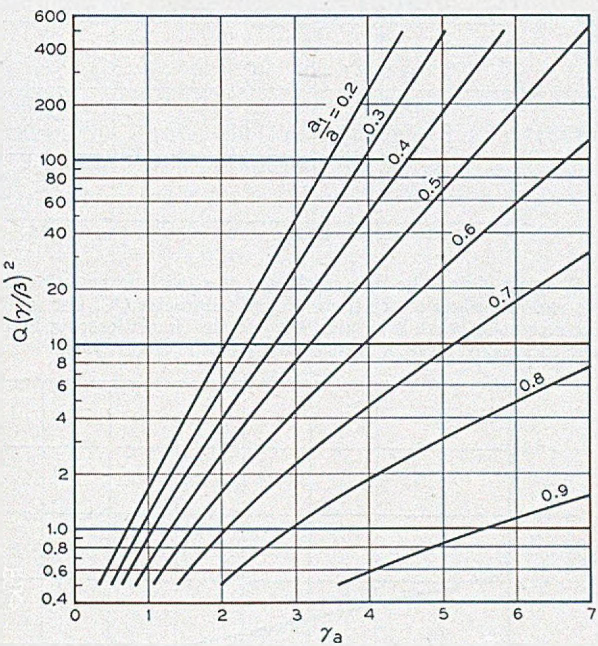


Fig. 8.12—Curves for obtaining Q for a helically conducting sheet and a hollow beam. The radius of the helically conducting sheet is a and that of the beam is a_1.

8.4 Differential Relations

It would be onerous to construct curves giving δ as a function of b for many values of attenuation and space charge. In some cases, however, useful information may be obtained by considering the effect of adding a small amount of attenuation when QC is large, or of seeing the effect of space charge when QC is small but the attenuation is large. We start with (7.13)

$$\delta^2 = \frac{1}{(-b + jd + j\delta)} - 4QC \tag{7.13}$$

Let us first differentiate (7.13) with respect to δ and d

$$2\delta\, d\delta = \frac{-j\, dd - j\, d\delta}{(-b + j\, d + j\delta)^2} \tag{8.29}$$

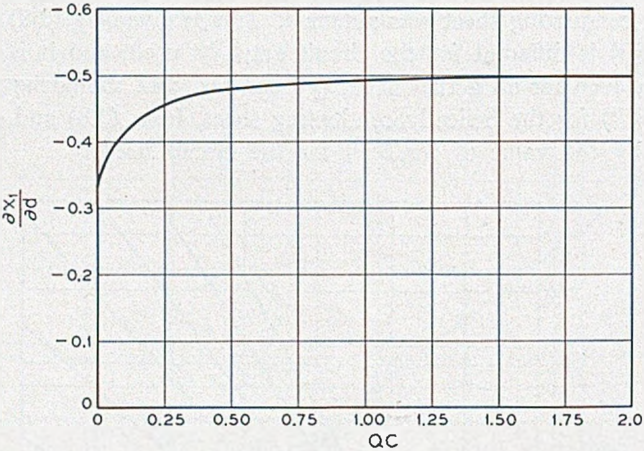


Fig. 8.13—A curve giving the rate of change of x_1 with attenuation parameter d for $d = 0$ and for various values of the space-charge parameter QC. For small values of QC the gain of the increasing wave is reduced by $\frac{1}{3}$ of the circuit loss; for large values of QC the gain of the increasing wave is reduced by $\frac{1}{2}$ of the circuit loss.

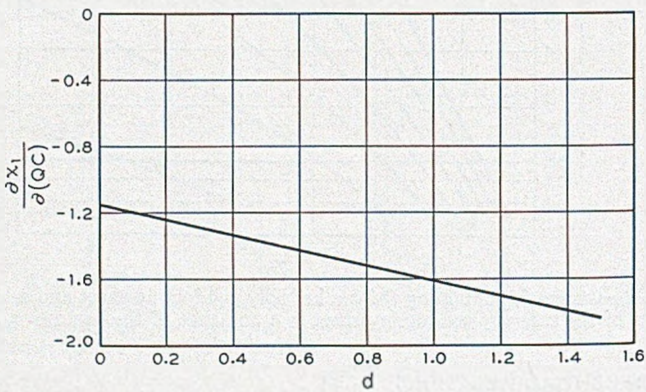


Fig. 8.14—A curve showing the variation of x_1 with QC for $QC = 0$ and for various values of the attenuation parameter d.

By using (7.13) we obtain

$$d\delta = \left(\frac{-j2\delta}{(\delta^2 + 4QC)^2} - 1 \right)^{-1} dd \tag{8.30}$$

If we allow d to be small, we can use the values of δ of Figs. 8.7–8.9 to plot the quantity

$$\mathrm{Re}(d\delta_1/dd) = dx_1/dd \tag{8.31}$$

vs. QC. In Fig. 8.13, this has been done for b chosen to make x_1 a maximum. We see that a small loss dd causes more reduction of gain as QC is increased (more space charge).

Let us now differentiate (7.13) with respect to QC

$$2\delta\, d\delta = \frac{-j\, d\delta}{(-b + j\, d + j\delta)^2} - 4\, d(QC) \tag{8.32}$$

By using (7.13) with $QC = 0$ we obtain

$$d\delta = \left(\frac{-4}{2\delta + j\delta^4}\right) d(QC) \tag{8.33}$$

In Fig. 8.14, $dx/d(QC)$ has been plotted vs. d for $b = 0$.

We see that the reduction of gain for a small amount of space charge becomes greater, the greater the loss is increased (d increased).

Both Fig. 8.13 and Fig. 8.14 indicate that for large values of QC or d the gain will be overestimated if space charge (QC) and loss (d) are considered separately.

CHAPTER IX

DISCONTINUITIES

Synopsis of Chapter

WE WANT TO KNOW the overall gain of traveling-wave tubes. So far, we have evaluated only the gain of the increasing wave, and we must find out how strong an increasing wave is set up when a voltage is applied to the circuit.

Beyond this, we may wish for some reason to break the circuit up into several sections having different parameters. For instance, it is desirable that a traveling-wave tube have more loss in the backward direction than it has gain in the forward direction. If this is not so, small mismatches will result either in oscillation or at least in the gain fluctuating violently with frequency. We have already seen in Chapter VIII the effect of a uniform loss in reducing the gain of the increasing wave. We need to know also the overall effect of short sections of loss in order to know how loss may best be introduced.

Such problems are treated in this chapter by matching boundary conditions at the points of discontinuity. It is assumed that there is no reflected wave at the discontinuity. This will be very nearly so, because the characteristic impedances of the waves differ little over the range of loss and velocity considered. Thus, the total voltages, a-c convection currents and the a-c velocities on the two sides of the point of discontinuity are set equal.

For instance, at the beginning of the circuit, where the unmodulated electron stream enters, the total a-c velocity and the total a-c convection current—that is, the sums of the convection currents and the velocities for the three waves—are set equal to zero, and the sum of the voltages for the three waves is set equal to the applied voltage.

For the case of no loss ($d = 0$) and an electron velocity equal to circuit phase velocity ($b = 0$) we find that the three waves are set up with equal voltages, each $\frac{1}{3}$ of the applied voltage. The voltage along the circuit will then be the sum of the voltages of the three waves, and the way in which the magnitude of this sum varies with distance along the circuit is shown in Fig. 9.1. Here CN measures distance from the beginning of the circuit and the amplitude relative to the applied voltage is measured in db.

The dashed curve represents the voltage of the increasing wave alone.

For large values of CN corresponding to large gains, the increasing wave predominates and we can neglect the effect of the other waves. This leads to the gain expression

$$G = A + BCN \text{ db}$$

Here BCN is the gain in db of the increasing wave and A measures its initial level with respect to the applied voltage.

In Fig. 9.2, A is plotted vs. b for several values of the loss parameter d. The fact that A goes to ∞ for $d = 0$ as b approaches $(3/2)\ (2)^{1/3}$ does not imply an infinite gain for, at this value of b, the gain of the increasing wave approaches zero and the voltage of the decreasing wave approaches the negative of that for the increasing wave.

Figure 9.3 shows how A varies with d for $b = 0$. Figure 9.4 shows how A varies with QC for $d = 0$ and for b chosen to give a maximum value of B (the greatest gain of the increasing wave).

Suppose that for $b = QC = 0$ the loss parameter is suddenly changed from zero to some finite value d. Suppose also that the increasing wave is very large compared with the other waves reaching the discontinuity. We can then calculate the ratio of the increasing wave just beyond the discontinuity to the increasing wave reaching the discontinuity. The solid line of Fig. 9.5 shows this ratio expressed in decibels. We see that the voltage of the increasing wave excited in the lossy section is less than the voltage of the incident increasing wave.

Now, suppose the waves travel on in the lossy section until the increasing wave again predominates. If the circuit is then made suddenly lossless, we find that the increasing wave excited in this lossless section will have a greater voltage than the increasing wave incident from the lossy section, as shown by the dashed curve of Fig. 9.5. This increase is almost as great as the loss in entering the lossy section. Imagine a tube with a long lossless section, a long lossy section and another long lossless section. We see that the gain of this tube will be less than that of a lossless tube of the same total length by about the reduction of the gain of the increasing wave in lossy section.

Suppose that the electromagnetic energy of the circuit is suddenly absorbed at a distance beyond the input measured by CN. This might be done by severing a helix and terminating the ends. The a-c velocity and convection current will be unaffected in passing the discontinuity, but the circuit voltage drops to zero. For $d = b = QC = 0$, Fig. 9.6 shows the ratio of V_1, the amplitude of the increasing wave beyond the break, to V, the amplitude the increasing wave would have had if there were no break. We see that for CN greater than about 0.2 the loss due to the break is not serious. For CN large (the break far from the input) the loss approaches 3.52 db.

Beyond such a break, the total voltage increases with CN as shown in Fig. 9.7, and from $CN = 0.2$ the circuit voltage is very nearly equal to the voltage of the increasing wave.

Often, for practical reasons loss is introduced over a considerable distance, sometimes by putting lossy material near to a helix. Suppose we use CN computed as if for a lossless section of circuit as a measure of length of the lossy section, and assume that the loss is great enough so that the circuit voltage (as opposed to that produced by space charge) can be taken as zero. Such a lossy section acts as a drift space. Suppose that an increasing wave only reaches this lossy section. The amplitude of the increasing wave excited beyond the lossy section in db with respect to the amplitude of the increasing wave reaching the lossy section is shown vs. CN, which measures the length of the lossy section, in Fig. 9.8.

9.1 General Boundary Conditions

We have already assumed that C is small, and when this is so the characteristic impedance of the various waves is near to the circuit characteristic impedance K. We will neglect any reflections caused by differences among the characteristic impedances of the various waves.

We will consider cases in which the circuit is terminated in the $+z$ direction, so as to give no backward wave. We will then be concerned with the 3 forward waves, for which δ has the values δ_1, δ_2, δ_3 and the waves represented by these values of δ have voltages V_1, V_2, V_3, electron velocities v_1, v_2, v_3 and convection currents i_1, i_2, i_3.

Let V, v, i be the total voltage, velocity and convection current at $z = 0$. Then we have

$$V_1 + V_2 + V_3 = V \tag{9.1}$$

and from (7.15) and (7.16),

$$\frac{V_1}{\delta_1} + \frac{V_2}{\delta_2} + \frac{V_3}{\delta_3} = (ju_0C/\eta)v \tag{9.2}$$

$$\frac{V_1}{\delta_1^2} + \frac{V_2}{\delta_2^2} + \frac{V_3}{\delta_3^2} = (-2V_0C^2/I_0)i \tag{9.3}$$

These equations yield, when solved,

$$V_1 = [V - (\delta_2 + \delta_3)(ju_0C/\eta)v + \delta_2\delta_3(-2V_0C^2/I_0)i] \cdot [(1 - \delta_2/\delta_1)(1 - \delta_3/\delta_1)]^{-1} \tag{9.4}$$

We can obtain the corresponding expressions for V_2 and V_3 simply by interchanging subscripts; to obtain V_2, for instance, we substitute subscript 2 for 1 and subscript 1 for 2 in (9.4).

9.2 Lossless Helix, Synchronous Velocity, No Space Change

Suppose we consider the case in which $b = d = Q = 0$, so that we have the values of δ obtained in Chapter II

$$\begin{aligned} \delta_1 &= e^{-j\pi/6} = \sqrt{3}/2 - j1/2 \\ \delta_2 &= e^{-j5\pi/6} = -\sqrt{3}/2 - j1/2 \\ \delta_3 &= e^{j\pi/2} = j \end{aligned} \tag{9.5}$$

Suppose we inject an unmodulated electron stream into the helix and apply a voltage V. The obvious thing is to say that, at $z = 0$, $v = i = 0$. It is not quite clear, however, that $v = 0$ at $z = 0$ (the beginning of the circuit). Whether or not there is a stray field, which will give an initial velocity modulation, depends on the type of circuit. Two things are true, however. For the small values of C usually encountered such a velocity modulation constitutes a small effect. Also, the fields of the first part of the helix act essentially to velocity modulate the electron stream, and hence a neglect of any small initial velocity modulation will be about equivalent to a small displacement of the origin.

If, then, we let $v = i = 0$ and use (9.4) we obtain

$$V_1 = V[(1 - \delta_2/\delta_1)(1 - \delta_3/\delta_1)]^{-1} \tag{9.6}$$

$$V_1 = V/3 \tag{9.7}$$

Similarly, we find that

$$V_2 = V_3 = V/3 \tag{9.8}$$

We have used V to denote the voltage at $z = 0$. Let V_z be the voltage at z. We have

$$\begin{aligned} V_z &= (V/3)e^{-j\beta_e z}(e^{j(1/2)\beta_e Cz+(\sqrt{3}/2)\beta_e Cz} + e^{j(1/2)\beta_e Cz-(\sqrt{3}/2)\beta_e Cz} + e^{j\beta_e Cz}) \\ V_z &= (V/3)e^{-j\beta_e(1-C)z}\ (1 + 2\cosh((\sqrt{3}/2)\beta eCz)e^{-j(3/2)\beta_e Cz}) \end{aligned} \tag{9.9}$$

From this we obtain

$$\begin{aligned} |V_z/V|^2 = (1/9)[1 &+ 4\cosh^2(\sqrt{3}/2)\beta_e Cz \\ &+ 4\cos(3/2)\beta_e Cz \cosh(\sqrt{3}/2)\beta_e Cz] \end{aligned} \tag{9.10}$$

We can express gain in db as $10 \log_{10} |V_z/V|^2$, and, in Fig. 9.1, gain in db is plotted vs CN, where N is the number of cycles.

We see that initially the voltage does not change with distance. This is natural, because the electron stream initially has no convection current, and hence cannot act on the circuit until it becomes bunched. Finally, of course, the increasing wave must predominate over the other two, and the slope of the line must be

$$B = 47.3/CN \tag{9.11}$$

The dashed line represents the increasing wave, which starts at $V_s/V = \frac{1}{3}$ (-9.54 db) and has the slope specified by (9.11). Thus, if we write for the increasing wave that gain G is

$$G = A + BCN \text{ db} \tag{9.12}$$

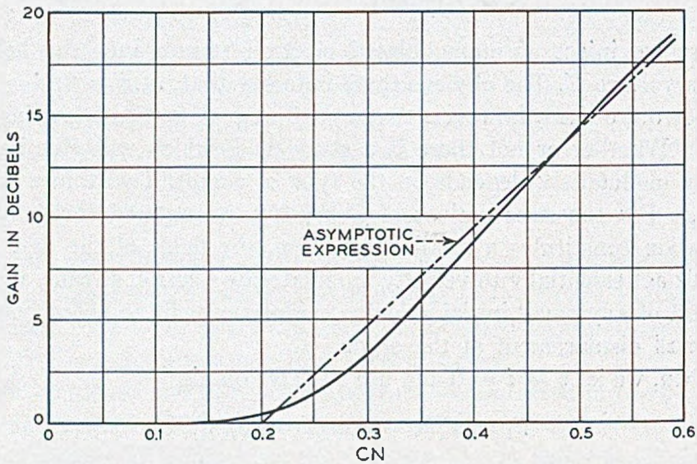


Fig. 9.1—How the signal level varies along a traveling-wave tube for the special case of zero loss and space charge and an electron velocity equal to the circuit phase velocity (solid curve). The dashed curve is the level of the increasing wave alone, which starts off with $\frac{1}{3}$ of the applied voltage, or at -9.54 *db*.

This is an asymptotic expression for the total voltage at large values of z, where $|V_1| \gg |V_2|, |V_3|$, and for $b = d = Q = 0$

$$\begin{aligned} A &= -\ 9.54 \text{ db} \\ B &= 47.3 \end{aligned} \tag{9.13}$$

We see that (9.11) is pretty good for $CN > .4$, and not too bad for $CN > .2$.

9.3 Loss in Helix

In Chapter VIII, curves were given for δ_1, δ_2, δ_3 vs. b for $QC = 0$ and for d, the loss parameter, equal to 0, 0.5 and 1. From the data from which these curves were derived one can calculate the initial loss parameter by means of (9.6)

$$A = 20 \log_{10} | V_1/V | \tag{9.14}$$

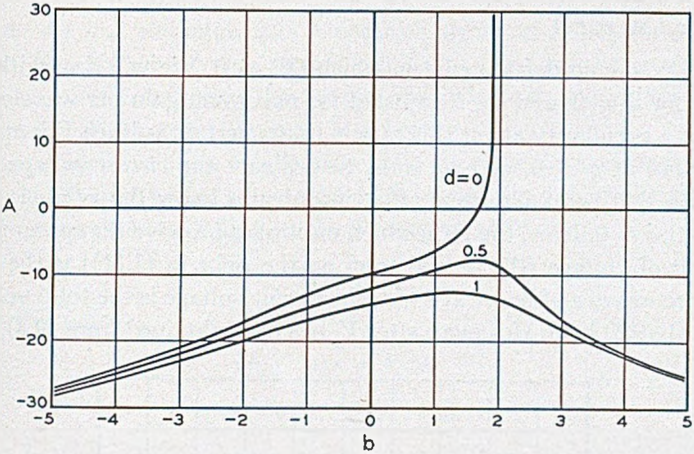


Fig. 9.2—When the gain is large we need consider the increasing wave only. Using this approximation, the gain in db is $A + BCN$ db. Here A is shown vs the velocity parameter b, several values of the attenuation parameter d, for no space charge ($QC = 0$).

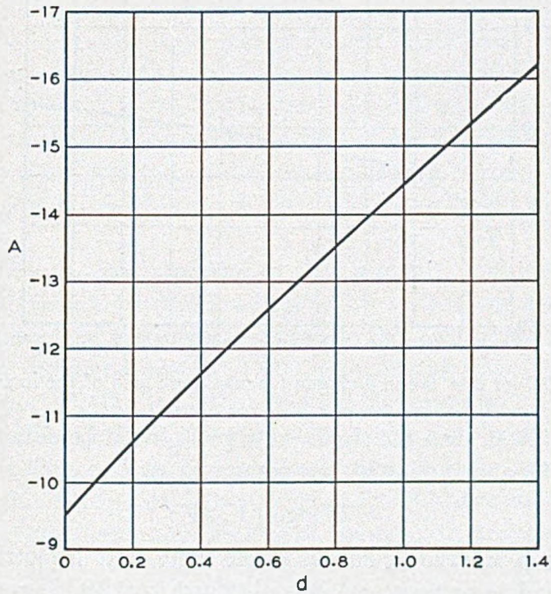


Fig. 9.3—A vs d for $b = 0$ and $QC = 0$.

In Fig. 9.2, A is plotted vs b for these three values of d.

It is perhaps of some interest to plot A vs d for $b = 0$ (the electron velocity equal to the phase velocity of the undisturbed wave). Such a plot is shown in Fig. 9.3.

9.4. Space Charge

We will now consider the case in which $QC \neq 0$. We will deal with this case only for $d = 0$, and for b adjusted for maximum gain per wavelength.

There is a peculiarity about this case in that a certain voltage V is applied to the *circuit* at $z = 0$, and we want to evaluate the *circuit* voltage associated with the increasing wave, V_{c1}, in order to know the gain.

At $z = 0$, $i = 0$. Now, the term which multiplies i to give the space-charge component of voltage (the second term on the right in (7.11)) is the same for all three waves and hence at $z = 0$ the circuit voltage is the total voltage. Thus, (9.1)–(9.3) hold. However, after V_1 has been obtained from (9.4), with

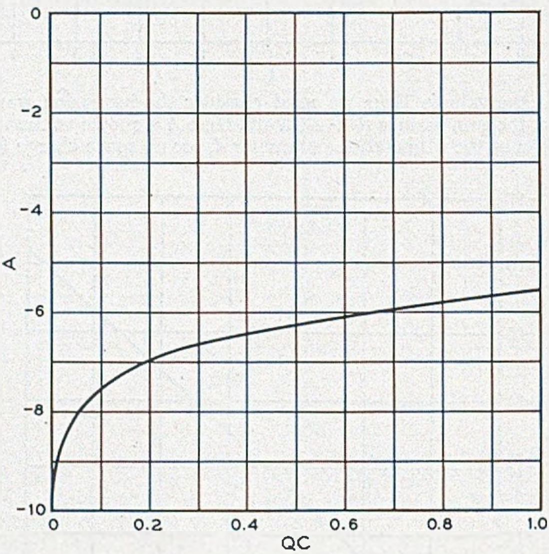


Fig. 9.4—A vs QC for $d = 0$ and b chosen for maximum gain of the increasing wave.

$V = V_1$, $v = i = 0$, then the circuit voltage V_{c1} must be obtained through the use of (7.14), and the initial loss parameter is

$$A = 20 \log_{10} | V_{c1}/V | \tag{9.15}$$

By using the appropriate values of δ, the same used in plotting Figs. 8.1 and 8.7–8.9, the loss parameter A was obtained from (9.15) and plotted vs QC in Fig. 9.4.

9.5 Change in Loss

We might think it undesirable in introducing loss to make the whole length of the helix lossy. For instance, we might expect the power output to be higher if the last part of the helix had low loss. Also, from Figs. 8.2

and 8.3 we see that the initial loss A becomes higher as d is increased. This is natural, because the electron stream can act to cause gain only after it is bunched, and if the initial section of the circuit is lossy, the signal decays before the stream becomes strongly bunched.

Let us consider a section of a lossless helix which is far enough from the input so that the increasing wave predominates and the total voltage V can be taken as that corresponding to the increasing wave

$$V = V_1 \tag{9.16}$$

Then, at this point

$$(ju_0C/\eta)v = V_1/\delta_1 \tag{9.17}$$

$$(-2V_0C^2/I_0)i = V_1/\delta_1^2 \,. \tag{9.18}$$

Here δ_1 is the value for $d = 0$ (and, we assume, $b = 0$). If we substitute the values from (9.16) in (9.4), and use in (9.4) the values of δ corresponding to $b = Q = 0$, $d \neq 0$, and call the value of V_1 we obtain V_1', we obtain the ratio of the initial amplitude of the increasing wave in the lossy section to the value of the increasing wave just to the left of the lossy section. Thus, the loss in the amplitude of the increasing wave in going from a lossless to a lossy section is $20 \log_{10} | V_1'/V_1 |$. This loss is plotted vs d in Fig. 8.5.

This loss is accounted for by the fact that $| i_1/V_1 |$ becomes larger as the loss parameter d is increased. Thus, the convection current injected into the lossy section is insufficient to go with the voltage, and the voltage must fall.

If we go from a lossy section ($d \neq 0$, $b = 0$) to a lossless section ($d = 0$, $b = 0$) we start with an excess of convection current and $| V_1' |$, the initial amplitude of the increasing wave to the right of the discontinuity is greater than the amplitude $| V_1 |$ of the increasing wave to the left. In Fig. 9.5, $20 \log_{10} | V_1'/V_1 |$ is plotted vs d for this case also.

We see that if we go from a lossless section to a lossy section, and if the lossy section is long enough so that the increasing wave predominates at the end of it, and if we go back to a lossless section at the end of it, the net loss and gain at the discontinuities almost compensate, and even for $d = 3$ the net discontinuity loss is less than 1 db. This does not consider the reduction of gain of the increasing wave in the lossy section.

9.6 Severed Helix

If the loss introduced is distributed over the length of the helix, the gain will decrease as the loss is increased (Fig. 8.5). If, however, the loss is distributed over a very short section, we easily see that as the loss is increased more and more, the gain must approach a constant value. The circuit will be in effect severed as far as the electromagnetic wave is concerned, and any excitation in the output will be due to the a-c velocity and convection current of the electron stream which crosses the lossy section.

We will first idealize the situation and assume that the helix is severed and by some means terminated looking in each direction, so that the voltage falls from a value V to a value 0 in zero distance, while v and i remain unchanged.

We will consider a case in which $b = d = Q = 0$, and in which a voltage

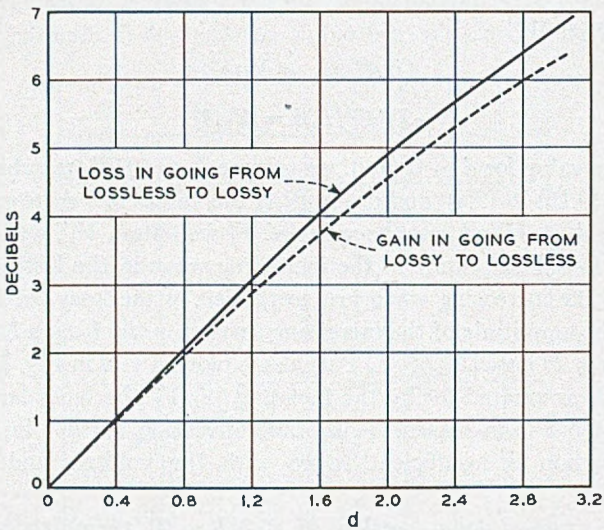


Fig. 9.5—Suppose that the circuit loss parameter changes suddenly with distance from 0 to d or from d to 0. Suppose there is an increasing wave only incident at the point of change. How large will the increasing wave beyond the point of change be? These curves tell ($b = QC = 0$).

V is applied to the helix N wavelengths before the cut. Then, just before the cut,

$$\begin{aligned} V_1 &= (V/3)e^{-j2\pi N}e^{2\pi NC\delta_1} \\ V_2 &= (V/3)e^{-j2\pi N}e^{2\pi NC\delta_2} \\ V_3 &= (V/3)e^{-j2\pi N}e^{2\pi NC\delta_3} \end{aligned} \tag{9.19}$$

and

$$\begin{aligned} (ju_0C/\eta)v_1 &= V_1/\delta_1 \\ (-2V_0C^3/I_0)i_1 &= V_1/\delta_1^2 \end{aligned} \tag{9.20}$$

etc.

Whence, just beyond the break which makes $V = 0$, V, v and i are

$$V = 0$$

$$(ju_0C/\eta)v = V_1/\delta_1 + V_2/\delta_2 + V_3/\delta_3 \tag{9.21}$$

$$(-2V_0C^3/I_0)i = V_1/\delta_1^2 + V_2/\delta_2^2 + V_3/\delta_3^2$$

Putting these values in (9.4), we can find V_1', the value of the increasing wave to the right of the break. The ratio of the magnitude of the increasing wave to the magnitude it would have if it were not for the break is then $|V_1'/V_1|$, and this ratio is plotted vs CN in Fig. 9.6, where N is the number of wavelengths in the first section.

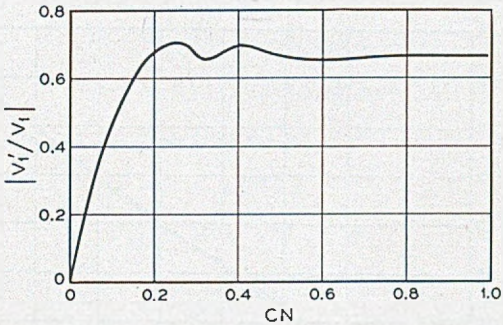


Fig. 9.6—Suppose the circuit is severed a distance measured by CN beyond the input, so that the voltage just beyond the break is zero. The ordinate is the ratio of the amplitude of the increasing wave beyond the break to that it would have had with an unbroken circuit ($b = QC = 0$).

We see that there will be least loss in severing the helix for CN equal to approximately $\frac{1}{4}$. From Fig. 9.1, we see that at $CN = \frac{1}{4}$ the voltage is just beginning to rise. In a typical 4,000 megacycle traveling-wave tube, CN is approximately unity for a 10 inch helix, so the loss should be put at least 2.5″ beyond the input. Putting the loss further on changes things little; asymptotically, $|V_1'/V|$ approaches $\frac{2}{3}$, or 3.52 db loss, for large values of CN (loss for from input).

It is of some interest to know how the voltage rises to the right of the cut. It was assumed that the cut was far from the point of excitation, so that only increasing wave of magnitude V_1 was present just to the left of the cut. The initial amplitudes of the three waves, V_1', V_2', V_3' to the right of the cut were computed and the magnitude of their sum plotted vs CN as it varies with distance to the right of the cut. The resulting curve, expressed in db with respect to the magnitude of the increasing wave V_1 just to the left of the cut, is shown in Fig. 9.7. Again, we see that at a distance $CN = \frac{1}{4}$ to the right of the cut the increasing wave (dashed straight line) predominates.

9.7 Severed Helix With Drift Space

In actually putting concentrated loss in a helix, the loss cannot be concentrated in a section of zero length for two reasons. In the first place, this is physically difficult if not impossible; in the second place it is desirable that the two halves of the helix be terminated in a reflectionless manner at the cut, and it is easiest to do this by tapering the loss. For instance, if the loss is put in by spraying aquadag (graphite in water) on ceramic rods supporting the helix, it is desirable to taper the loss coating at the ends of the lossy section.

Perhaps the best reasonably simple approximation we can make to such a lossy section is one in which the section starts far enough from the input

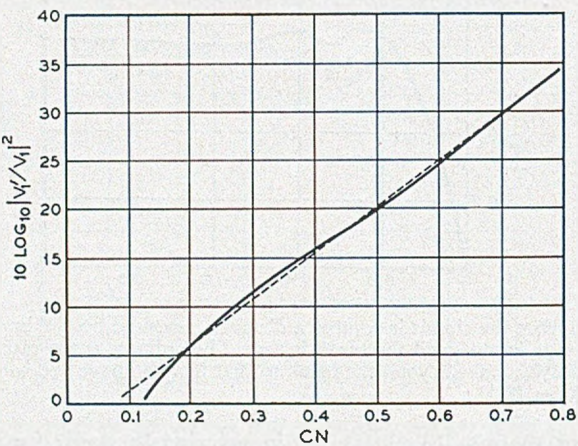


Fig. 9.7—Suppose that the circuit is severed and an increasing wave only is incident at the break. How does the signal build up beyond the break? The solid curve shows ($b = QC = 0$). 0 db is the level of the incident increasing wave.

so that at the beginning of the lossy section only an increasing wave is present. In the lossy section CN long we will consider that the loss completely shorts out the circuit, so that (8.28) holds. Thus, in the lossy section we will have only two values of δ, which we will call δ_I and δ_{II}.

$$\delta_I = jk \tag{9.21}$$

$$\delta_{II} = -jk \tag{9.22}$$

$$k = 2\sqrt{QC} \tag{9.23}$$

Let V_I and V_{II} be the voltages of the waves corresponding to δ_I and δ_{II} at the beginning of the lossy section. Let δ_1, δ_2, δ_3 be the values of δ to the left and right of the lossy section. Let V_1 be the amplitude of the increasing wave just to the left of the lossy section. Then, by equating velocities and convection currents at the start of the lossy section, we obtain

$$V_1/\delta_1 = V_I/\delta_I + V_{II}/\delta_{II} \tag{9.24}$$

and, from (9.21) and (9.22)

$$V_1/\delta_1 = (-j/k)(V_I - V_{II}) \tag{9.25}$$

Similarly

$$\begin{aligned} V_1/\delta_1^2 &= V_I/\delta_{II}^2 + V_{II}\,\delta_{II}^2 \\ V_1/\delta_1^2 &= -(1/k^2)(V_I + V_{II}) \end{aligned} \tag{9.26}$$

So that

$$V_I = j(V_1/2)(k/\delta_1)(jk/\delta_1 + 1) \tag{9.27}$$

$$V_{II} = j(V_1/2)(k/\delta_1)(jk/\delta_1 - 1) \tag{9.28}$$

At the output of the lossy section we have the voltages V_I' and V_{II}'

$$V_I' = V_I e^{-j2\pi N} e^{-j2\pi kCN} \tag{9.29}$$

$$V_{II}' = V_{II} e^{-j2\pi N} e^{-j2\pi kCN} \tag{9.30}$$

Thus, at the end of the lossy section we have

$$V = V_I' + V_{II}' \tag{9.31}$$

$$\begin{aligned} (ju_0C/\eta)v &= V_I'/\delta_I + V_{II}'/\delta_{II} \\ (ju_0C/\eta)v &= (-j/k)(V_I' - V_{II}') \end{aligned} \tag{9.32}$$

and similarly

$$(-2V_0C^2/I_0)i = (-1/k^2)(V_I' + V_{II}') \tag{9.33}$$

From (9.27) and (9.28) we see that

$$V_I' + V_{II}' = -(k/\delta_1)[+(k/\delta_1)\cos 2\pi kCN + \sin 2\pi kCN]V_1e^{-j2\pi N} \tag{9.34}$$

$$V_I' - V_{II}' = j(k/\delta_1)[-(k/\delta_1)\sin 2\pi kCN + \cos 2\pi kCN]V_1e^{-j2\pi N} \tag{9.35}$$

Whence

$$V = -(k/\delta_1)[+(k/\delta_1)\cos 2\pi kCN + \sin 2\pi kCN]V_1e^{-j2\pi N} \tag{9.36}$$

$$(ju_0C/\eta)v = (1/\delta_1)[-(k/\delta_1)\sin 2\pi kCN + \cos 2\pi kCN]V_1e^{-j2\pi N} \tag{9.37}$$

$$(-2V_0C^2/I_0)i = (1/\delta_1)[(1/\delta_1)\cos 2\pi kCN + (1/k)\sin 2\pi kCN]V_1e^{-j2\pi N} \tag{9.38}$$

These can be used in connection with (9.4) in obtaining V_1', the value of V_1 just beyond the lossy section; that is, the amplitude of the component of increasing wave just beyond the lossy section.

In typical traveling-wave tubes the lossy section usually has a length such that CN is $\frac{1}{4}$ or less. In Fig. 9.8 the loss in db in going through the lossy section, $20 \log_{10} | V_1'/V_1 |$, has been plotted vs. CN for $QC = 0$, .25, .5 for the range $CN = 0$ to $CN = .5$.

We see that, for low space charge, increasing the length of a drift space increases the loss. For higher space charge it may either increase or decrease the loss. It is not clear that the periodic behavior characteristic of the curves for $QC = 0.5$ and 1, for instance, will obtain for a drift space with tapered loss at each end. The calculations may also be considerably in error for broad electron beams (γa large). The electric field pattern in the helix differs

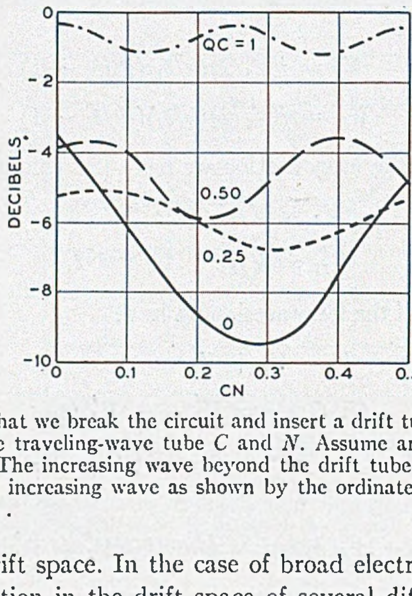


Fig. 9.8—Suppose that we break the circuit and insert a drift tube of length measured by CN in terms of the traveling-wave tube C and N. Assume an increasing wave only before the drift tube. The increasing wave beyond the drift tube will have a level with respect to the incident increasing wave as shown by the ordinate. Here $d = 0$ and b is chosen to maximize x_1.

from that in the drift space. In the case of broad electron beams this may result in the excitation in the drift space of several different space charge waves having different field patterns and different propagation constants.

A suggestion has been made that the introduction of loss itself has a bad effect. The only thing that affects the electrons is an electric field. Unpublished measurements made by Cutler mode by moving a probe along a helix indicate that in typical short high-loss sections the electric field of the helix is essentially zero. Hence, except for a short distance at the ends, such lossy sections should act simply as drift spaces.

9.8 Overall Behavior of Tubes

The material of Chapters VIII and IX is useful in designing traveling-wave tubes. Prediction of the performance of a given tube over a wide range of voltage and current is quite a different matter. For instance, in order to predict gain for voltage or current ranges for which the gain is small, the three waves must be taken into account. As current is varied, the loss parameter d varies, and this means different x's and y's must be computed for different currents. Finally, at high currents, the space-charge parameter Q must be taken into account. In all, a computation of tube behavior under a variety of conditions is an extensive job.

Fortunately, for useful tubes operating as intended, the gain is high. When this is so, the gain can be calculated quite accurately by asymptotic relations. Such an overall calculation of the gain of a helix-type tube with distributed loss is summarized in Appendix VII.

CHAPTER X

NOISE FIGURE

Synopsis of Chapter

Because there is no treatment of the behavior at high frequencies of an electron flow with a Maxwellian distribution of velocities, one might think there could be no very satisfactory calculation of the noise figure of traveling-wave tubes. Various approximate calculations can be made, and two of these will be discussed here. Experience indicates that the second and more elaborate of these is fairly well founded. In each case, an approximation is made in which the actual multi-velocity electron current is replaced by a current of electrons having a single velocity at a given point but having a mean square fluctuation of velocity or current equal to a mean square fluctuation characteristic of the multi-velocity flow.

In one sort of calculation, it is assumed that the noise is due to a current fluctuation equal to that of shot noise (equation (10.1)) in the current entering the circuit. For zero loss, an electron velocity equal to the phase velocity of the circuit and no space charge, this leads to an expression for noise figure (10.5), which contains a term proportional to beam voltage V_0 times the gain parameter C. One can, if he wishes, add a space-charge noise reduction factor multiplying the term 80 V_0C. This approach indicates that the voltage and the gain per wavelength should be reduced in order to improve the noise figure.

In another approach, equations applying to single-valued-velocity flow between parallel planes are assumed to apply from the cathode to the circuit, and the fluctuations in the actual multi-velocity stream are represented by fluctuations in current and velocity at the cathode surface. It is found that for space-charge-limited emission the current fluctuation has no effect, and so all the noise can be expressed in terms of fluctuations in the velocity of emission of electrons.

For a special case, that of a gun with an anode at circuit potential V_0, a cathode-anode transit angle θ_1, and an anode-circuit transit angle θ_2, an expression for noise figure (10.28) is obtained. This expression can be rewritten in terms of a parameter L which is a function of P

$$F = 1 + (\tfrac{1}{2})(4 - \pi)(T_c/T)(1/C)L$$

$$P = (\theta_1 - \theta_2)C$$

Formally, F can be minimized by choosing the proper value of P. In Fig. 10.3, the minimum value of L, L_m, is plotted vs. the velocity parameter b for zero loss and zero space charge ($d = QC = 0$). The corresponding value of P, P_m, is also shown.

P is a function of the cathode-anode transit angle θ_1, which cannot be varied without changing the current density and hence C, and of anode-circuit transit angle θ_2, which can be given any value. Thus, P can be made very small if one wishes, but it cannot be made indefinitely large, and it is not clear that P can always be made equal to P_m. On the other hand, these expressions have been worked out for a rather limited case: an anode potential equal to circuit potential, and no a-c space charge. It is possible that an optimization with respect to gun anode potential and space charge parameter QC would predict even lower noise figures, and perhaps at attainable values of the parameters.

In an actual tube there are, of course, sources of noise which have been neglected. Experimental work indicates that partition noise is very important and must be taken into account.

10.1 Shot Noise in the Injected Current

A stream of electrons emitted from a temperature-limited cathode has a mean square fluctuation in convection current $\overline{i_s^2}$

$$\overline{i_s^2} = 2eI_0B_0 \tag{10.1}$$

Here e is the charge on an electron, I_0 is the average or d-c current and B is the bandwidth in which the frequencies of the current components whose mean square value is $\overline{i_s^2}$ lie. Suppose this fluctuation in the beam current of a traveling-wave tube were the sole cause of an increasing wave ($V = v = 0$). Then, from (9.4) the mean square value of that increasing wave,, $\overline{V_{1s}^2}$, would be

$$\overline{V_{1s}^2} = (8eBV_0^2C^4/I_0) \mid \delta_2\delta_3 \mid^2 \mid (1 - \delta_2/\delta_1)(1 - \delta_3/\delta_1) \mid^{-2} \tag{10.2}$$

Now, suppose we have an additional noise source: thermal noise voltage applied to the circuit. If the helix is matched to a source of temperature T, the thermal noise power P_t drawn from the source is

$$P_t = kTB \tag{10.3}$$

Here k is Boltzman's constant, T is temperature in degrees Kelvin and, as before, B is bandwidth in cycles. If K_t is the longitudinal impedance of the circuit the mean square noise voltage $\overline{V_t^2}$ associated with the circuit will be

$$\overline{V_t^2} = kTBK_t \tag{10.4}$$

and the component of increasing wave excited by this voltage, $\overline{V_{1t}^2}$, will be, from (9.4),

$$\overline{V_{1t}^2} = kTBK_t \mid (1 - \delta_2/\delta_1)(1 - \delta_3/\delta_1) \mid^{-2} \tag{10.5}$$

The noise figure of an amplifier is defined as the ratio of the total noise output power to the noise output power attributable to thermal noise at the input alone. We will regard the mean-square value of the initial voltage V_1 of the increasing wave as a measure of noise output. This will be substantially true if the signal becomes large prior to the introduction of further noise. For example, it will be substantially true in a tube with a severed helix if the helix is cut at a point where the increasing wave has grown large compared with the original fluctuations in the electron stream which set it up.

Under these circumstances, the noise figure F will be given by

$$F = (\overline{V_{1s}^2} + \overline{V_{1t}^2})/(\overline{V_{1t}^2})$$

$$F = 1 + (e/kT)(8V_0^2C^4/I_0K_t) \mid \delta_2\delta_3 \mid^2 \tag{10.3}$$

Now we have from Chapter II that

$$C^3 = I_0K_t/4V_0$$

whence

$$F = 1 + 2(eV_0/kT)C \mid \delta_2\delta_3 \mid^2 \tag{10.4}$$

The standard reference temperature is 290°K. Let us assume $b = d = QC = 0$. For this case we have found $\mid \delta_2 \mid = \mid \delta_3 \mid = 1$. Thus, for these assumptions we find

$$F = 1 + 80V_0C \tag{10.5}$$

A typical value of V_0 is 1,600 volts; a typical value of C is .025. For these values

$$F = 3{,}201$$

In db this is a noise of 35 db.

This is not far from the noise figure of traveling-wave tubes when the cathode temperature is lowered so as to give temperature-limited emission. The noise figure of traveling-wave tubes in which the cathode is at normal operating temperature and is active, so that emission is limited by space-charge, can be considerably lower. In endeavoring to calculate the noise figure for space-charge-limited electron flow from the cathode we must proceed in a somewhat different manner.

10.2 The Diode Equations

Llewellyn and Peterson[1] have published a set of equations governing the behavior of parallel plane diodes with a single-valued electron velocity. They sum up the behavior of such a diode in terms of nine coefficients A^*–I^*, in the following equations

$$V_b - V_a = A^* I + B^* q_a + C^* v_a \tag{10.6}$$

$$q_b = D^* I + E^* q_a + F^* v_a \tag{10.7}$$

$$v_b = G^* I + H^* q_a + I^* v_a \tag{10.8}$$

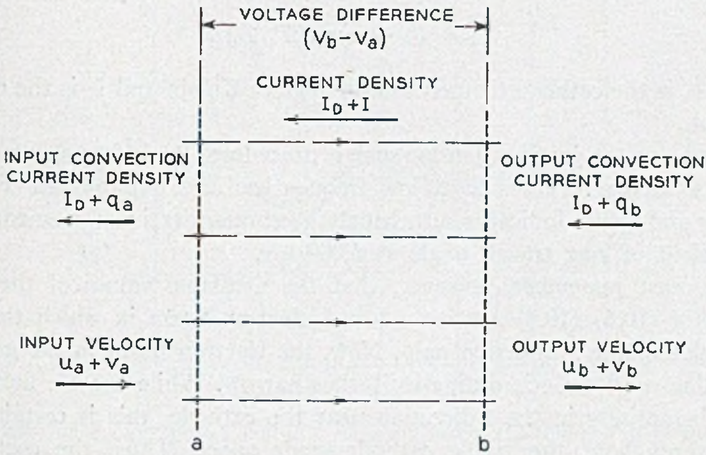


Fig. 10.1—Parallel electron flow between two planes *a* and *b* normal to the flow, showing the currents, velocities and voltages.

These equations and the values of the various coefficients in terms of current, electron velocity and transit angle are given in Appendix V. The diode structure to which they apply is indicated in Fig. 10.1. Electrons enter normal to the left plane and pass out at the right plane. The various quantities involved are transit angle between the two planes and:

I_0	d-c current density to left
I	a-c current density to left
q_a	a-c convection current density to left at input plane *a*
q_b	a-c convection current density to left at output plane *b*
u_a	d-c velocity to right at plane *a*
u_b	d-c velocity to right at plane *b*
v_a	a-c velocity to right at plane *a*
v_b	a-c velocity to right at plane *b*
V_b-V_a	a-c potential difference between plane *b* and plane *a*

[1] F. B. Llewellyn and L. C. Peterson, "Vacuum Tube Networks," *Proc. I.R.E.*, Vol. 32, pp. 144–166, March, 1944.

We will notice that I and the q's are current *densities* and that, contrary to the convention we have used, they are taken as positive to the left. Thus, if the area is σ, we would write the output convection current; as

$$i = -\sigma q_b$$

where q_b is the convection current density used in (10.6)–(10.8).

Peterson has used (10.6)–(10.8) in calculating noise figure by replacing the actual multi-velocity flow from the cathode by a single-velocity flow with the same mean square fluctuation in velocity, namely,[2]

$$\overline{v_t^2} = (4 - \pi)\eta\,(kT_c/I_0)B \tag{10.9}$$

Here T_c is the cathode temperature in degrees Kelvin and I_0 is the cathode current.

Whatever the justification for such a procedure, Rack[3] has shown that it gives a satisfactory result at low frequencies, and unpublished work by Cutler and Quate indicates surprisingly good quantitative agreement under conditions of long transit angle at 4,000 mc.

We must remember, however, that the available values of the coefficients of (10.6)–(10.8) are for a broad electron beam in which there are a-c fields in the z direction only. Now, the electron beam in the gun of a traveling-wave tube is ordinarily rather narrow. While the a-c fields may be substantially in the z-direction near the cathode, this is certainly not true throughout the whole cathode-anode space. Thus, the coefficients used in (10.6)–(10.8) are certainly somewhat in error when applied to traveling-wave tube guns.

Various plausible efforts can be made to amend this situation, as, by saying that the latter part of the beam in the gun acts as a drift region in which the electron velocities are not changed by space-charge fields. However, when one starts such patching, he does not know where to stop. In the light of available knowledge, it seems best to use the coefficients as they stand for the cathode-anode region of the gun.

Let us then consider the electron gun of the traveling-wave tube to form a space-charge limited diode which is short-circuited at high frequencies. If we assume complete space charge (space-charge limited emission) and take the electron velocity at the cathode to be zero, we find that the quantities multiplying q_a in (10.6)–(10.8) are zero.

$$B^* = E^* = H^* = 0^* \tag{10.10}$$

[2] L. C. Peterson, "Space-Charge and Transit-Time Effects on Signal and Noise in Microwave Tetrodes," *Proc. I.R.E.*, Vol. 35, pp. 1264–1272, November, 1947.
[3] A. J. Rack, "Effect of Space Charge and Transit Time on the Shot Noise in Diodes," *Bell System Technical Journal*, Vol. 17, pp. 592–619, October, 1938.

Accordingly, the magnitude of the noise convection current at the cathode does not matter. If we assume that the gun is a short-circuited diode as far as r-f goes

$$V_b - V_a = 0 \tag{10.11}$$

Then from (10.6), (10.10) and (10.11) we obtain

$$I = -\frac{C^*}{A^*} v_a \tag{10.12}$$

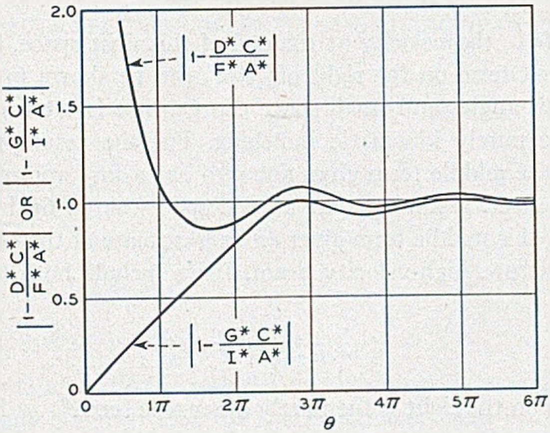


Fig. 10.2—Some expressions useful in noise calculations, showing how they approach unity at large transit angles.

Accordingly, from (10.7) and (10.8) we obtain

$$q_b = \left(1 - \frac{D^*C^*}{F^*A^*}\right) F^* v_a \tag{10.13}$$

$$v_b = \left(1 - \frac{G^*C^*}{I^*A^*}\right) I^* v_a \tag{10.14}$$

In Fig. 10.2, $|1 - D^*C^*/F^*A^*|$ and $|1 - G^*C^*/I^*A^*|$ are plotted vs θ, the transit angle. We see that for transit angles greater than about 3π these quantities differ negligibly from unity, and we may write

$$q_b = F^* v_a \tag{10.15}$$

$$v_b = I^* v_a \tag{10.16}$$

More specifically, we find

$$q_b = \frac{v_a I_0 \beta_1 e^{-\beta_1}}{u_b} \tag{10.17}$$

$$v_b = -v_a e^{-\beta_1} \tag{10.18}$$

Here β_1 is j times the transit angle in radians from cathode to anode. For v_a we use a velocity fluctuation with the mean-square value given by (10.9).

Suppose now that there is a constant-potential drift space following the diode anode, of length β_2/j in radians. If we apply (10.6)–(10.8) and assume that the space-charge is small and the transit angle long, we find that q_b', the value of q_b at the end of this drift space, is given in terms of q_a' and v_a', the values at the beginning of this drift space, by

$$q_b' = (q_a' + (I_0/u_b)\beta_2 v_a')e^{-\beta_2} \tag{10.19}$$

The case of v_b', the velocity at the end of this drift space, is a little different. The first term on the right of (10.8) can be shown to be negligible for long transit angles and small space charge. The last term on the right represents the purely kinematic bunching. For the assumption of small space charge the middle term gives not zero but a first approximation of a space-charge effect, assuming that all the space-charge field acts longitudinally. Thus, this middle term gives an overestimate of the effect of space-charge in a narrow, high-velocity beam. If we include both terms, we obtain

$$v_b' = H_2^* q_a' + e^{-\beta_2} v_a' \tag{10.20}$$

Here the term on the right is the purely kinematic term.*

Now, the current from the gun is assumed to go into the drift space, so that q_a' is q_b from (10.17) and v_a' is v_a from (10.18). The *d-c* velocity at the gun anode and throughout the drift space are both given by u_b. If we make these substitutions in (10.19) and (10.20) we obtain

$$q_b' = (I_0/u_b)(\beta_1 - \beta_2)e^{-(\beta_1+\beta_2)} v_a \tag{10.21}$$

$$v_b' = -\left(2\frac{\beta_1}{\beta_2} + 1\right)e^{-(\beta_1+\beta_2)}\, v_a \tag{10.22}$$

The term $2\beta_1/\beta_2$ in (10.22) is the "space-charge" term. We will in the following analysis omit this, making the same sort of error we do in neglecting space charge in the traveling-wave section of the tube. If space charge in the drift space is to be taken into account, it is much better to proceed as in 9.7.

From the drift-space the current goes into the helix. It is now necessary to change to the notation we have used in connection with the traveling-wave tube. The chief difference is that we have taken currents as positive to the right, but allowed I_0 to be the d-c current to the left. If i and v are

* The first term has been written as shown because it is easiest to use the small space-charge value of H^* for the drift region (H_2^*) in connection with the space-charge limited value of F^* for the cathode-anode region rather than in connection with (10.17).

our a-c convection current and velocity at the beginning of the helix, and I_0 and u_0 the d-c beam current and velocity, and σ the area of the beam,

$$
\begin{aligned}
i &= -\sigma q_b' \\
v &= v_b \\
I_0 &= \sigma I_0 \\
u_0 &= u_0
\end{aligned}
\tag{10.23}
$$

In addition, we will use transit angles θ_1 and θ_2 in place of β_1 and β_2

$$
\begin{aligned}
\beta_1 &= j\theta_1 \\
\beta_2 &= j\theta_2
\end{aligned}
\tag{10.24}
$$

We then obtain from (10.21) and (10.22)

$$q = -j(I_0/u_0)(\theta_1 - \theta_2)e^{-j(\theta_1+\theta_2)}v_a \tag{10.25}$$

$$v = -e^{-j(\theta_1+\theta_2)}v_a \tag{10.26}$$

10.3 Overall Noise Figure

We are now in a position to use (9.4) in obtaining the overall noise figure. We have already assumed that the space-charge is small in the drift space between the gun anode and the helix ($QC = 0$). If we continue to assume this in connection with (9.4), the only voltage is the helix voltage and for the noise caused by the velocity fluctuation at the cathode, v_a, $V = 0$ at the beginning of the helix. Thus, the mean square initial noise voltage of the increasing wave, $\overline{V_{1s}^2}$, will be, from (10.21), (10.22), (9.4) and (10.9),

$$
\begin{aligned}
\overline{V_{1s}^2} = (2(4 - \pi)kT_cCBV_0/I_0)| \delta_2\delta_3(\theta_1 - \theta_2)C + (\delta_2 + \delta_3) |^2 \\
| (1 - \delta_2/\delta_1)(1 - \delta_3/\delta_1) |^{-2}
\end{aligned}
\tag{10.27}
$$

As before, we have, from the thermal noise input to the helix

$$\overline{V_{1t}^2} = kTBK_t| (1 - \delta_2/\delta_1)(1 - \delta_3/\delta_1) |^{-2} \tag{10.5}$$

and the noise figure becomes

$$F = 1 + \overline{V_{1s}^2}/\overline{V_{1t}^2}$$

$$F = 1 + (1/2)(4 - \pi)(T_c/T)(1/C)| \delta_2\delta_3(\theta_1 - \theta_2)C + (\delta_2 + \delta_3) |^2 \tag{10.28}$$

Here use has been made of the fact that

$$C = K_tI/4V_0$$

Let us investigate this for the case $b = d = 0$ (we have already assumed $QC = 0$). In this case

$$\delta_2 = \sqrt{3}/2 - j1/2$$

$$\delta_3 = j$$

and we obtain

$$F = 1 + (1/2)(4 - \pi)(T_c/T)(1/C)\left| (P/2 - \sqrt{3}/2) - j(\sqrt{3}P/2 - 1/2) \right|^2 \tag{10.29}$$

$$P = (\theta_1 - \theta_2)C \tag{10.30}$$

For a given gun transit-angle θ_1, the parameter P can be given values ranging from $\theta_1 C$ to large negative values by increasing the drift angle θ_2 between the gun anode and the beginning of the helix.

We see that

$$F = 1 + (1/2)(4 - \pi)(T_c/T)(1/C)(P^2 - \sqrt{3}P + 1) \tag{10.31}$$

The minimum value of $(P^2 - \sqrt{3}P + 1)$ occurs when

$$P = \sqrt{3}/2 \tag{10.32}$$

if the product of the gun transit angle and C is large enough, this can be attained. The corresponding value of $(P^2 - \sqrt{3}P + 1)$ is $\frac{1}{4}$, and the corresponding noise figure is

$$F = 1 + (1/2)(1 - \pi/4)(T_c/T)(1/C) \tag{10.33}$$

A typical value for T_c is 1020°K, and for a reference temperature of 290°K,

$$T_c/T = 3.5$$

A typical value of C is .025. For these values

$$F = 17$$

or a noise figure of 12 db.

Let us consider cases for no attenuation or space-charge but for other electron velocities. In this case we write, as before

$$\delta_2 = x_2 + jy_2$$

$$\delta_3 = x_3 + jy_3$$

Let us write, for convenience,

$$L = \left| \delta_2\delta_3 P + \delta_1 + \delta_2 \right|^2 \tag{10.34}$$

Then we find that

$$L = [(x_2x_3)^2 + (y_2y_3)^2 + (x_2y_3)^2 + (x_3y_2)^2]P^2 + 2[x_3(y_2^2 + x_2^2) + x_2(x_3^2 + y_3^2)]P + (x_2 + x_3)^2 + (y_2 + y_3)^2 \qquad (10.35)$$

This has a minimum value for $P = P_m$

$$P_m = \frac{-[x_3(x_2^2 + y_2^2) + x_2(x_3^2 + y_3^2)]}{(x_2x_3)^2 + (y_2y_3)^2 + (x_2y_3)^2 + (x_3y_2)^2} \qquad (10.36)$$

We note that, as we are not dealing with the increasing wave, x_2 and x_3

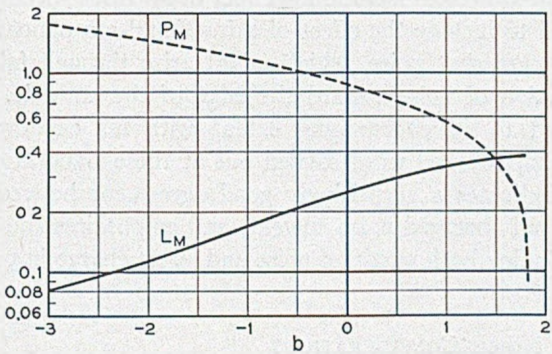


Fig. 10.3—According to the theory presented, the overall noise figure of a tube with a lossless helix and no space charge is proportional to L. Here we have a minimum value of L_m, minimized with respect to P, which is dependent on gun transit angle, and also the corresponding value of P, P_m. According to this curve, the optimum noise figure should be lowest for low electron velocities (low values of b). It may, however, be impossible to make P equal to P_m.

must be either negative or zero, and hence P_m is always positive. For no space-charge and no attenuation, x_3 is zero for all values of b and

$$P_m = \frac{-x_2}{y_2^2 + x_2^2} \qquad (10.37)$$

From (10.36) and (10.35), the minimum value of L, L_m, is

$$L_m = (x_2 + x_3)^2 + (y_2 + y_3)^2 - \frac{[x_3(y_2^2 + x_2^2) + x_2(x_3^2 + y_3^2)]^2}{(x_2x_3)^2 + (y_2y_3)^2 + (x_2y_3)^2 + (x_3y_2)^2} \qquad (10.38)$$

When $x_3 = 0$, as in (10.37)

$$L_m = x_2^2 + y_2^2 + 2y_2y_3 + \frac{y_2^2y_3^2}{x_2^2 + y_2^2} \qquad (10.39)$$

In Fig. 10.3, P_m and L_m are plotted vs b for no attenuation ($d = 0$). We see that P_m becomes very small as b approaches $(3/2)2^{1/3}$, the value at which the increasing wave disappears.

If space charge is to be taken into account, it should be taken into account both in the drift space between anode and helix and in the helix itself. In the helix we can express the effect of space-charge by means of the parameter QC and boundary conditions can be fitted as in Chapter IX. The drift space can be dealt with as in Section 9.7 of Chapter IX. The inclusion of the effect of space-charge by this means will of course considerably complicate the analysis, especially if $b \neq 0$.

While working with Field at Stanford, Dr. C. F. Quate extended the theory presented here to include the effect of all three waves in the case of low gain, and to include the effect of a fractional component of beam current having pure shot noise, which might arise through failure of space-charge reduction of noise toward the edge of the cathode. His extended theory agreed to an encouraging extent with his experimental results. Subsequent unpublished work carried out at these Laboratories by Cutler and Quate indicates a surprisingly good agreement between calculations of this sort and observed noise current, and emphasizes the importance of properly including both partition noise and space charge in predicting noise figure.

10.4 Other Noise Considerations

Space-charge reduction of noise is a cooperative phenomenon of the whole electron beam. If some electrons are eliminated, as by a grid, additional "partition" noise is introduced. Peterson shows how to take this into account.[2]

An electron may be ineffective in a traveling-wave tube not only by being lost but by entering the circuit near the axis where the r-f field is weak rather than near the edge where the r-f field is high. Partition noise arises because sidewise components of thermal velocity cause a fluctuation in the amount of current striking a grid or other intercepting circuit. If such sidewise components of velocity appreciably alter electron position in the helix, a noise analogous to partition noise may arise even if no electrons actually strike the helix. Such a noise will also occur if the "counteracting pulses" of low-charge density which are assumed to smooth out the electron flow are broad transverse to the beam.

These considerations lead to some maxims in connection with low-noise traveling-wave tubes: (1) do not allow electrons to be intercepted by various electrodes (2) if practical, make sure that $I_0(\beta r)$ is reasonably constant over the beam, and/or (3) provide a very strong magnetic focusing field, so that electrons cannot move appreciably transversely.

10.5 Noise in Transverse-Field Tubes

Traveling-wave tubes can be made in which there is no longitudinal field component at the nominal beam position. One can argue that, if a narrow, well-collimated beam is used in such a tube, the noise current in the beam can induce little noise signal in the circuit (none at all for a beam of zero thickness with no sidewise motion). Thus, the idea of using a transverse-field tube as a low-noise tube is attractive. So far, no experimental results on such tubes have been announced.

A brief analysis of transverse-field tubes is given in Chapter XIII.

CHAPTER XI

BACKWARD WAVES

WE NOTED IN CHAPTER IV that, in filter-type circuits, there is an infinite number of spatial harmonics which travel in both directions. Usually, in a tube which is designed to make use of a given forward component the velocity of other forward components is enough different from that of the component chosen to avoid any appreciable interaction with the electron stream. It may well be, however, that a backward-traveling component has almost the same speed as a forward-traveling component.

Suppose, for instance, that a tube is designed to make use of a given forward-traveling component of a forward wave. Suppose that there is a forward-traveling component of a backward wave, and this forward-traveling component is also near synchronism with the electrons. Does this mean that under these circumstances both the backward-traveling and the forward-traveling waves will be amplified?

The question is essentially that of the interaction of an electron stream with a circuit in which the phase velocity is in step with the electrons but the group velocity and the energy flow are in a direction contrary to that of electron motion.

We can most easily evaluate such a situation by considering a distributed circuit for which this is true. Such a circuit is shown in Fig. 11.1. Here the series reactance X per unit length is negative as compared with the more usual circuit of Fig. 11.2. In the circuit of Fig. 11.2, the phase shift is 0° per section at zero frequency and assumes positive values as the frequency is increased. In the circuit of Fig. 11.1 the phase shift is $-180°$ per section at a lower cutoff frequency and approaches 0° per section as the frequency approaches infinity.

Suppose we consider the equations of Chapter II. In (2.9) we chose the sign of X in such a manner as to make the series reactance positive, as in Fig. 11.2, rather than negative, as in Fig. 11.1. All the other equations apply equally well to either circuit. Thus, for the circuit of Fig. 11.1, we have, instead of (2.10),

$$V = \frac{+\Gamma\Gamma_1 i}{(\Gamma^2 - \Gamma_1^2)} \qquad (11.1)$$

The sign is changed in the circuit equation relating the convection current and the voltage. Similarly, we can modify the equations of Chapter VII, (7.9) and (7.12), by changing the sign of the left-hand side. From Chapter VIII, the equation for a lossless circuit with no space charge is

$$\delta^2(\delta + jb) = -j \tag{8.1}$$

The corresponding modification is to change the sign preceding δ^2, giving

$$\delta^2(\delta + jb) = +j \tag{11.2}$$

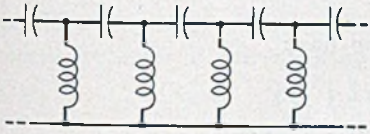


Fig. 11.1

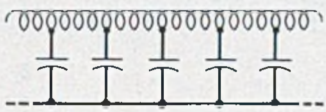


Fig. 11.2

Fig. 11.1—A circuit with a negative phase velocity. The electrons can be in synchronism with the field only if they travel in a direction opposite to that of electromagnetic energy flow.

Fig. 11.2—A circuit with a positive phase velocity.

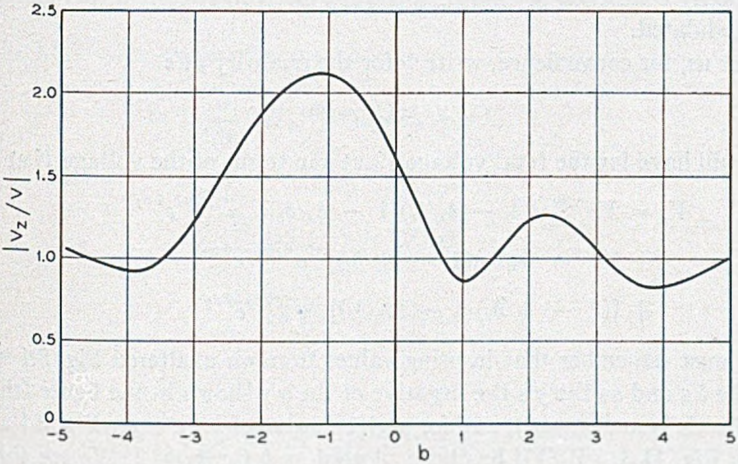


Fig. 11.3—Suppose we have a tube with a circuit such as that of Fig. 11.1, in which the circuit energy is really flowing in the opposite direction from the electron motion. Here, for $QC = d = 0$, we have the ratio of the magnitude of the voltage V_z a distance z from the point of injection of electrons to the magnitude of the voltage V at the point of injection of electrons. V_z is really the input voltage, and there will be gain at values of b for which $|V_z/V| < 1$.

In (11.2), b and δ have the usual meaning in terms of electron velocity and propagation constant.

Now consider the equation

$$\delta^2(\delta - jk) = j \tag{11.3}$$

Equations (11.2) and (8.1) apply to different systems. We have solutions of (8.1) and we want solutions of (11.2). We see that a solution of (11.2) is a solution of (11.3) for $k = -b$. We see that a solution of (11.3) is the conjugate of a solution of (8.1) if we put b in (8.1) equal to k in (11.3). Thus, a solution of (11.2) is the conjugate of a solution of (8.1) in which b in (8.1) is made the negative of the value of b for which it is desired to solve (11.2).

We can use the solutions of Fig. 8.1 in connection with the circuit of Fig. 11.1 in the following way: wherever in Fig. 8.1 we see b, we write in instead $-b$, and wherever we see y_1, y_2 or y_3 we write in instead $-y_1$, $-y_2$ or $-y_3$.

Thus, for synchronous velocity, we have

$$\delta_1 = \sqrt{3}/2 + j\tfrac{1}{2}$$

$$\delta_2 = -\sqrt{3}/2 + j\tfrac{1}{2}$$

$$\delta_3 = -j$$

We can determine what will happen in a physical case only by fitting boundary conditions so that at $z = 0$ the electron stream, as it must, enters unmodulated.

Let us, for convenience, write Φ for the quantity βCz

$$\beta Cz = \Phi \tag{11.4}$$

We will have for the total voltage V_z at z in terms of the voltage V at $z = 0$

$$\begin{aligned} V_z = Ve^{-j\beta z}([(1 - \delta_2/\delta_1)(1 - \delta_3/\delta_1)]^{-1}e^{-j\Phi y_1}e^{\Phi x_1} \\ + [(1 - \delta_3/\delta_2)(1 - \delta_1/\delta_2)]^{-1}e^{-j\Phi y_2}e^{\Phi x_2} \\ + [(1 - \delta_1/\delta_3)(1 - \delta_2/\delta_3)]^{-1}e^{-j\Phi y_3}e^{\Phi x_3}) \end{aligned} \tag{11.5}$$

We must remember that in using values from an unaltered Fig. 8.1 we use in the δ's and as the y's the negative of the y's shown in the figure (the sign of the x's is unchanged), and for a given value of b we enter Fig. 8.1 at $-b$.

In Fig. 11.3, $|V_z/V|$ has been plotted vs b for $\Phi = 2$. We see that, for several values of b, $|V_z|$ (the input voltage) is less than $|V|$ (the output voltage) and hence there can be "backward" gain.

We note that as Φ is made very large, the wave which increases with increasing Φ will eventually predominate, and $|V_z|$ will be greater than $|V|$. "Backward gain" occurs not through a "growing wave" but rather through a sort of interference between wave components, as exhibited in Fig. 11.3.

Fig. 11.3 is for a lossless circuit; the presence of circuit attenuation would alter the situation somewhat.

APPENDIX IV

EVALUATION OF SPACE—CHARGE PARAMETER Q

Consider the system consisting of a conducting cylinder of radius a and an internal cylinder of current of radius a_1 with a current

$$ie^{j\omega t}e^{-\Gamma z}. \tag{1}$$

Let subscript 1 refer to inside and 2 to outside. We will assume magnetic fields of the form

$$H_{\varphi 1} = AI_1(\gamma r) \tag{2}$$

$$H_{\varphi 2} = BI_1(\gamma r) + CK_1(\gamma r) \tag{3}$$

From Maxwell's equations we have,

$$\frac{\partial}{\partial r}(rH_\varphi) = j\omega\epsilon rE_z + rJ_z \tag{4}$$

Now

$$\frac{\partial}{\partial z}(zI_1(z)) = zI_0(z) \tag{5}$$

$$\frac{\partial}{\partial z}(zK_1(z)) = -zK_0(z) \tag{6}$$

Hence

$$E_{z1} = \frac{-j\gamma}{\omega\epsilon}AI_0(\gamma r) \tag{7}$$

$$E_{z2} = \frac{-j\gamma}{\omega\epsilon}(BI_0(\gamma r) - CK_0(\gamma r)) \tag{8}$$

at $r = a$, $E_{z2} = 0$

$$C = B\frac{I_0(\gamma a)}{K_0(\gamma a)} \tag{9}$$

at $r = a_1$, $E_{z1} = E_{z2}$

$$\begin{aligned} AI_0(\gamma a_1) &= B\left(I_0(\gamma a_1) - \frac{I_0(\gamma a)}{K_0(\gamma a)}K_0(\gamma a_1)\right) \\ A &= B\left(1 - \frac{I_0(\gamma a)}{K_0(\gamma a)}\frac{K_0(\gamma a_1)}{I_0(\gamma a_1)}\right) \end{aligned} \tag{10}$$

In going across boundary, we integrate (4) over the infinitesimal radial distance which the current is assumed to occupy

$$\begin{aligned} rdH_\varphi &= rJdr \\ 2\pi rJdr &= i \\ rjdr &= \frac{i}{2\pi} \end{aligned} \tag{11}$$

Thus

$$dH_\varphi = \frac{i}{2\pi r} = \frac{i}{2\pi a_1} = (H_{\varphi 2} - H_{\varphi 1})_{a_1} \tag{12}$$

$$B\left[I_1(\gamma a_1) + \frac{I_0(\gamma a)}{K_0(\gamma a)} K_1(\gamma a_1) - I_1(\gamma a_1)\left(1 - \frac{I_0(\gamma a)K_0(\gamma a_1)}{K_0(\gamma a)I_0(\gamma a_1)}\right)\right] = \frac{i}{2\pi a_1}$$

$$B = \frac{i}{2\pi a_1}\left[\frac{I_0(\gamma a)}{K_0(\gamma a)} K_1(\gamma a_1) + \frac{I_0(\gamma a)}{K_0(\gamma a)} \frac{K_0(\gamma a_1)}{I_0(\gamma a_1)} I_1(\gamma a_1)\right]^{-1}$$

$$B = \frac{i}{2\pi a_1} \frac{K_0(\gamma a)}{I_0(\gamma a)I_1(\gamma a_1)}\left[\frac{K_1(\gamma a_1)}{I_1(\gamma a_1)} + \frac{K_0(\gamma a_1)}{I_0(\gamma a_1)}\right]^{-1} \tag{13}$$

at $r = a_1$

$$\begin{aligned} E_{z1} = E_{z2} = {} & \left(\frac{-j\gamma}{\omega\epsilon}\right)\left(\frac{i}{2\pi a_1}\right)\frac{K_0(\gamma a)}{I_0(\gamma a)}\frac{I_0(\gamma a_1)}{I_1(\gamma a_1)} \\ & \left(1 - \frac{I_0(\gamma a)}{K_0(\gamma a)}\frac{K_0(\gamma a_1)}{I_0(\gamma a_1)}\right)\left[\frac{K_1(\gamma a_1)}{I_1(\gamma a_1)} + \frac{K_0(\gamma a_1)}{I_0(\gamma a_1)}\right]^{-1} \end{aligned} \tag{14}$$

Now

$$\frac{1}{\omega\epsilon} = \frac{\sqrt{\mu/\epsilon}}{\beta_0} = \frac{377}{\beta_0} \tag{15}$$

Hence

$$\begin{aligned} i\beta V &= E_z = j\,\frac{\gamma^2}{\beta_0}\, I_0^2(\gamma a_1)G(\gamma a, \gamma a_1)i \\ V &= \left(\frac{\gamma}{\beta_0}\right)\left(\frac{\gamma}{\beta}\right) I_0^2(\gamma a_1)G(\gamma a, \gamma a_1)q \end{aligned} \tag{16}$$

$$G(\gamma a, \gamma a_1) = 60\left[\frac{K_0(\gamma a_1)}{I_0(\gamma a_1)} - \frac{K_0(\gamma a)}{I_0(\lambda a)}\right] \tag{17}$$

In obtaining this form, use was made of the fact that

$$K_1(z)I_0(z) + K_0(z)I_1(z) = \frac{1}{z}$$

Now

$$Q = \frac{\beta}{\omega C_1(E^2/\beta^2 P)} \tag{18}$$

where $(E^2/\beta^2 P)$ is the value of this quantity at $r = a_1$. In order to evaluate Q we note that

$$V = -\frac{j\Gamma}{\omega C_1} i = \frac{-j(j\beta)}{\omega C_1} i$$

$$V = \frac{\beta}{\omega C_1} i$$

$$\frac{\beta}{\omega C_1} = \frac{V}{i} = \left(\frac{\gamma}{\beta_0}\right)\left(\frac{\gamma}{\beta}\right) I_0^2(\gamma a_1) G(\gamma a, \gamma a_1) \tag{20}$$

$$\frac{\beta}{\omega C_1} = \left(\frac{\beta}{\beta_0}\right)\left(\frac{\gamma}{\beta}\right)^2 I_0^2(\gamma a_1) G(\gamma a, \gamma a_1)$$

On the axis, $(E^2/\beta^2 P)$ has a value $(E^2/\beta^2 P)_0$

$$(E^2/\beta^2 P)_0 = \left(\frac{\beta}{\beta_0}\right)\left(\frac{\gamma}{\beta}\right)^4 F^3(a) \tag{21}$$

At a radius a_1

$$(E^2/\beta^2 P) = \left(\frac{\beta}{\beta_0}\right)\left(\frac{\gamma}{\beta}\right)^4 F^3(\gamma a) I_0^2(\gamma a_1) \tag{22}$$

Hence

$$Q(\gamma/\beta)^2 = \frac{G(\gamma a, \gamma a_1)}{F^3(\gamma a)} \tag{23}$$

APPENDIX V

DIODE EQUATIONS

FROM LLEWELLYN AND PETERSON

These apply to electrons injected into a space between two planes a and b normal to the x direction. Plan b is in the $+x$ direction from plane a. Current density I and convection current q are positive in the $-x$ direction. The d-c velocities u_a, u_b and the a-c velocities v_a, v_b are in the $+x$ direction. T is the transit time. The notation in this appendix should not be confused with that used in other parts of this book. It was felt that it would be confusing to change the notation in Llewellyn's and Peterson's[1] well-known equations.

TABLE I

ELECTRONICS EQUATIONS

Numerics Employed:

$$\eta = 10^7 \frac{e}{m} = 1.77 \times 10^{15}, \qquad \epsilon = 1/(36\pi \times 10^{11}) \frac{\eta}{\epsilon} \doteq 2 \times 10^{28}$$

Direct-Current Equations:

Potential-velocity: $\eta V_D = (1/2)u^2$ (1)

$$\left.\begin{array}{l} \text{Space-charge-factor definition: } \zeta = 3(1 - T_0/T) \\ \text{Distance: } x = (1 - \zeta/3)(u_a + u_a)T/2 \\ \text{Current density: } (\eta/\epsilon)I_D = (u_a + u_b)2\zeta/T^2 \end{array}\right\} \quad (2)$$

Space-charge ratio: $I_D/I_m = (9/4)\zeta(1 - \zeta/3)^2$ (3)

Limiting-current density:

$$I_m = \frac{2.33}{10^6} \frac{(\sqrt{V_{Da}} + \sqrt{V_{Db}})^3}{x^2} \quad (4)$$

Alternating-Current Equations:

Symbols employed:

$\beta = i\theta, \quad \theta = \omega T, \quad i = \sqrt{-1}$

[1] F. B. Llewellyn and L. C. Peterson "Vacuum Tube Networks," *Proc. I.R.E.*, vol. 32, pp. 144–166, March, 1944.

$$P = 1 - e^{-\beta} - \beta e^{-\beta} \doteq \frac{\beta^2}{2} - \frac{\beta^3}{3} + \frac{\beta^4}{8} \cdots$$

$$Q = 1 - e^{-\beta} \doteq \beta - \frac{\beta^2}{2} + \frac{\beta^3}{6} - \frac{\beta^4}{24} \cdots$$

$$S = 2 - 2e^{-\beta} - \beta - \beta e^{-\beta} \doteq -\frac{\beta^3}{6} + \frac{\beta^4}{12} - \frac{\beta^5}{40} + \frac{\beta^6}{180}$$

General equations for alternating current

q = alternating conduction-current density

v = alternating velocity

$$\left.\begin{aligned} V_b - V_a &= A^*I + B^*q_a + C^*v_a \\ q_b &= D^*I + E^*q_a + F^*v_a \\ v_b &= G^*I + H^*q_a + I^*v_a \end{aligned}\right\} \quad (5)$$

TABLE II

VALUES OF ALTERNATING-CURRENT COEFFICIENTS

$$A^* = \frac{1}{\epsilon}\, u_a + u_b \frac{T^2}{2}\frac{1}{\beta} \left[1 - \frac{\zeta}{3}\left(1 - \frac{12S}{\beta^3}\right)\right]$$

$$B^* = \frac{1}{\epsilon}\frac{T^2}{\beta^3}\,[u_a(P - \beta Q) - u_b P + \zeta(u_a + u_b)P]$$

$$C^* = -\frac{1}{\eta}\, 2\zeta(u_a + u_b)\frac{P}{\beta^2}$$

$$D^* = 2\zeta\,\frac{(u_a + u_b)}{u_b}\frac{P}{\beta^2}$$

$$E^* = \frac{1}{u_b}\,[u_b - \zeta(u_a + u_b)]e^{-\beta}$$

$$F^* = \frac{\epsilon}{\eta}\frac{2\zeta}{T^2}\frac{(u_a + u_b)}{u_b}\,\beta e^{-\beta}$$

$$G^* = -\frac{\eta}{\epsilon}\frac{T^2}{\beta^3}\frac{1}{u_b}\,[u_b(P - \beta Q) - u_a P + \zeta(u_a + u_b)P]$$

$$H^* = -\frac{\eta}{\epsilon}\frac{T^2}{2}\frac{(u_a + u_b)}{u_b}(1 - \zeta)\frac{e^{-\beta}}{\beta}$$

$$I^* = \frac{1}{u_b}\,[u_a - \zeta(u_a + u_b)]e^{-\beta}$$

Complete space-charge, $\zeta = 1$.

$$A^* = \frac{1}{\epsilon}\,(u_a + u_b)\frac{T^2}{3\beta}\left(1 + \frac{6S}{\beta^3}\right)$$

$$B^* = \frac{1}{\epsilon}\frac{T^2}{\beta^3}\, u_a(2P - \beta Q)$$

$$C^* = -\frac{2}{\eta}(u_a + u_b)\frac{P}{\beta^2}$$

$$D^* = 2\frac{(u_a + u_b)}{(u_b)}\frac{P}{\beta^2}$$

$$E^* = -\frac{u_a}{u_b}e^{-\beta}$$

$$F^* = \frac{\epsilon}{\eta}\frac{2}{T^2}\frac{(u_a + u_b)}{(u_b)}\beta e^{-\beta}$$

$$G^* = -\frac{\eta}{\epsilon}\frac{T^2}{\beta^3}(2P - \beta Q)$$

$$H^* = 0$$

$$I^* = -e^{-\beta}$$

APPENDIX VI

EVALUATION OF IMPEDANCE AND Q FOR THIN AND SOLID BEAMS[1]

Let us first consider a thin beam whose breadth is small enough so that the field acting on the electrons is essentially constant. The normal mode solutions obtained in Chapters VI and VII apply only to this case. The more practical situation of a thick beam will be considered later. The normal mode method consists of simultaneously solving two equations, one relating the r-f field produced on the circuit by an impressed r-f current from the electron stream and the other relating r-f current produced in the electron stream by an impressed r-f field from the circuit.

We have the circuit equation

$$E = -\left[\frac{\Gamma^2\Gamma_0 K}{\Gamma^2 - \Gamma_0^2} + \frac{2jQK\Gamma^2}{\beta_e}\right] i \tag{1}$$

and the electronic equation

$$i = \frac{j\beta_e}{(j\beta_e - \Gamma)^2}\,\frac{I_0}{2V_0}\,E. \tag{2}$$

The solution of these two equations gives Γ in terms of Γ_0, K, and Q, which must be evaluated separately for the particular circuit being considered.

The field solution is obtained by solving the field equations in various regions and appropriately matching at the boundaries. For a hollow beam of electrons of radius b traveling in the z direction inside a helix of radius a and pitch angle ψ, the matching consists of finding the admittances $\left(\frac{H_\varphi}{E_z}\right)$ inside and outside the beam and setting the difference equal to the admittance of the beam. Thus the admittance just outside the beam for an idealized helix will be[2]

$$Y_0 = \frac{H_{\varphi 0}}{E_{z0}} = j\,\frac{\omega\epsilon}{\gamma}\,\frac{I_1(\gamma b) - \delta K_1(\gamma b)}{I_0(\gamma b) + \delta K_0(\gamma b)}, \tag{3}$$

[1] This appendix is taken from R. C. Fletcher, "Helix Parameters in Traveling-Wave Tube Theory," *Proc. I.R.E.*, Vol. 38, pp. 413–417 (1950).

[2] L. J. Chu and J. D. Jackson, "Field Theory of Traveling-Wave Tubes," *I.R.E., Proc.*, Vol. 36, pp. 853–863, July, 1948.

O. E. H. Rydbeck, "Theory of the Traveling-Wave Tube," *Ericsson Technics*, No. 46 pp. 3–18, 1948.

where

$$\delta = \frac{1}{K_0^2(\gamma a)} \left(\left(\frac{\beta_0 a \cot \Psi}{\gamma a} \right)^2 I_1(\gamma a) K_1(\gamma a) - I_0(\gamma a) K_0(\gamma a) \right),$$

$$\beta_0^2 = \omega^2 \mu \epsilon,$$

and

$$\gamma^2 = -\Gamma^2 - \beta_0^2 .$$

(The I's and K's are modified Bessel functions). The admittance inside the beam is

$$Y_i = \frac{H_{\varphi i}}{E_{zi}} = \frac{j\omega\epsilon}{\gamma} \frac{I_1(\gamma b)}{I_0(\gamma b)} . \tag{4}$$

Boundary conditions require that $E_{z0} = E_{zi} = E_z$ and $H_{z0} - H_{zi} = \dfrac{i}{2\pi b}$. Combining the boundary conditions, we see that

$$Y_0 - Y_i = \frac{1}{2\pi b} \frac{i}{E_z} , \tag{5}$$

where the ratio of $\dfrac{i}{E_z}$ is given by (2). Thus the field method gives two equations which are equivalent to the circuit and electronic equations of the normal mode method.

A6.1 Normal Mode Parameters for Thin Beam

The constants appearing in eq. (1) can be evaluated by equating the circuit equation (1) to the circuit equation (5). Thus if $Y_c = Y_0 - Y_i$,

$$-\frac{\Gamma^2 \Gamma_0 K}{\Gamma^2 - \Gamma_0^2} - \frac{2jQK\Gamma^2}{\beta_e} = + \frac{1}{2\pi b Y_c} . \tag{6}$$

The constants can be obtained by expanding each side of eq. (6) in terms of the zero and pole occurring in the vicinity of Γ_0. Thus if γ_0 and γ_p are the zero and pole of Y_c, respectively,

$$Y_c \simeq -(\gamma_p - \gamma_0) \left(\frac{\partial Y_c}{\partial \gamma} \right)_{\gamma = \gamma_0} \left(\frac{\gamma - \gamma_0}{\gamma - \gamma_p} \right), \tag{7}$$

and the two sides of eq. (6) will be equivalent if

$$\Gamma_0^2 = -\gamma_0^2 - \beta_0^2 , \tag{8}$$

$$\frac{2Q}{\beta_e} = \left(1 + \frac{\beta_0^2}{\gamma_0^2} \right)^{-1/2} \frac{\gamma_0}{\gamma_p^2 - \gamma_0^2} , \tag{9}$$

and

$$\frac{1}{K} = -j\pi b\gamma_0^2\left(1 + \frac{\beta_0^2}{\gamma_0^2}\right)^{3/2}\left(\frac{\partial Y_c}{\partial Y}\right)_{\gamma=\gamma_0}. \tag{10}$$

γ_0 and γ_p can be obtained from eqs. (3) and (4) through the implicit equations

$$(\beta a \cot \Psi)^2 = (\gamma_0 a)^2 \frac{I_0(\gamma_0 a)K_0(\gamma_0 a)}{I_1(\gamma_0 a)K_1(\gamma_0 a)}, \tag{11}$$

$$\begin{aligned}\frac{I_0(\gamma_p b)}{K_0(\gamma_p b)} = &-\frac{1}{K_0^2(\gamma_p a)} \\ &\cdot\left[\left(\frac{\beta_0 a \cot \Psi}{\gamma_p a}\right)^2 I_1(\gamma_p a)K_1(\gamma_p a) - I_0(\gamma_p a)K_0(\gamma_p a)\right],\end{aligned} \tag{12}$$

and $1/K$ is found to be

$$\begin{aligned}\frac{1}{K} = \pi\sqrt{\frac{\epsilon}{\mu}}\left(1 + \frac{\beta_0^2}{\gamma_0^2}\right)^{3/2} \frac{\beta_0^2}{I_0^2(\gamma_0 b)}\frac{I_0(\gamma_0 a)}{K_0(\gamma_0 a)}&\left[\frac{I_1(\gamma_0 a)}{I_0(\gamma_0 a)} - \frac{I_0(\gamma_0 a)}{I_1(\gamma_0 a)}\right. \\ &\left.+\frac{K_0(\gamma_0 a)}{K_1(\gamma_0 a)} - \frac{K_1(\gamma_0 a)}{K_0(\gamma_0 a)} + \frac{4}{\gamma_0 a}\right].\end{aligned} \tag{13}$$

The equations for γ_0 and K are the same as those given by Appendix II, evaluated by solving the field equations for the helix without electrons present. The evaluation of γ_p, and thus Q, represents a new contribution. Values of $Q\frac{\gamma_0}{\beta_e}\left(1 + \frac{\beta_0^2}{\gamma_0^2}\right)^{-1/2}$ are plotted in Fig. A6.1 as a function of $\gamma_0 a$ for various ratios of b/a. (It should be noted that for most practical applications the factor $\frac{\gamma_0}{\beta_e}\left(1 + \frac{\beta_0^2}{\gamma_0^2}\right)^{-1/2}$ is very close to unity, so that the ordinate is practically the value of Q itself.)

Appendix IV gives a method for estimating Q based on the solution of the field equations for a conductor replacing the helix and considering the resultant field to be $-\frac{2jKQ\Gamma^2}{\beta_e}i$. This estimate of Q is plotted as the dashed lines of Fig. A6.1.

A6.2 Thick Beam Case

For an electron beam which entirely fills the space out to the radius b, the electronic equations of both the normal mode method and the field method are altered in such a way as to considerably complicate the solution. In order to find a solution for this case some simplifying assumptions must be made. A convenient type of assumption is to replace the thick beam by an "equivalent" thin beam, for which the solutions have already been worked out.

Two beams will be equivalent if the value of $\frac{H_\varphi}{E_z}$ is the same outside the beams, since the matching to the circuit depends only on this admittance.

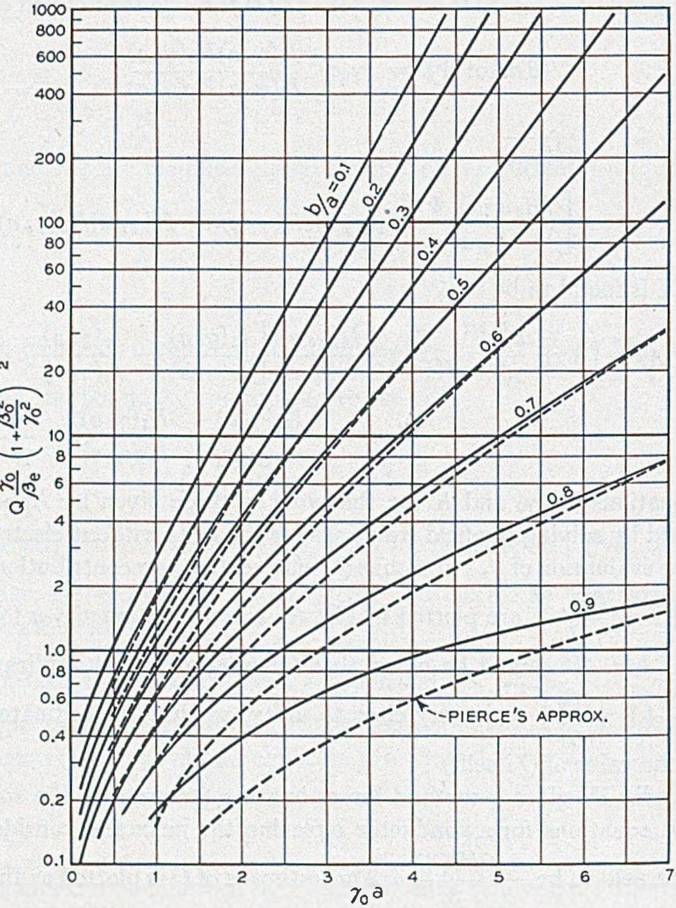


Fig. A6.1—Passive mode parameter Q for a hollow beam of electrons of radius b inside a helix of radius a and natural propagation constant γ_0. The solid line was obtained by equating the circuit equation of the normal mode method, which defines Q, with a corresponding circuit equation found from the field theory method. The dashed line was obtained in Appendix IV from a solution of the field equations for a conductor replacing the helix.

The problem, then, of making a thin beam the equivalent of a thick beam is the problem of arranging the position and current of a thin beam to give the same admittance at the radius b of the thick beam. This is of course impossible for all values of γ. It is desirable therefore that the admittances be the same close to the complex values of γ which will eventually solve the equations.

The solution of the field equations for the solid beam yields the value for $\frac{H_\varphi^{(1)}}{E_z}$ at the radius b as

$$\frac{H_\varphi}{E_z} = \frac{j\omega\epsilon}{\gamma} \frac{nI_1(n\gamma b)}{I_0(n\gamma b)}, \tag{14}$$

where

$$n^2 = 1 + \frac{1}{\beta_0}\sqrt{\frac{\mu}{\epsilon}}\,\frac{\beta_e I_0}{2\pi b^2 V_0}\,\frac{1}{(j\beta_e - \Gamma)^2}. \tag{15}$$

Thus the electronic equation for the solid beam which must be solved simultaneously with the circuit equation (given above by either the normal mode approximation or the field solution) must be

$$Y_e = \frac{H_\varphi}{E_z} - Y_i = \frac{j\omega\epsilon b}{\gamma b}\left[\frac{nI_1(n\gamma b)}{I_0(n\gamma b)} - \frac{I_1(\gamma b)}{I_0(\gamma b)}\right]. \tag{16}$$

Complex roots for γ will be expected in the vicinity of real values of γ for which $Y_e \approx Y_c$ and $\frac{dY_e}{d\gamma} \approx \frac{dY_c}{d\gamma}$. By plotting Y_e and Y_c vs. real values of γ, it is found that the two curves become tangent close to the value of γ for which $n = 0$, using typical operating conditions (Fig. A6.2). Our procedure for choosing a hollow beam equivalent of the solid beam, then, will be to equate the values of Y_e and $\frac{dY_e}{d\gamma}$ at $n = 0$. This will give us two equations from which to solve for the electron beam diameter and d-c current for the equivalent hollow beam.

If the hollow beam is placed at the radius sb with a current of tI_0, the value of $\frac{H_\varphi}{E_z}$ at the radius b gives the value for Y_{eH} as

$$\begin{aligned} Y_{eH} = \left(\frac{H_\varphi}{E_z}\right)_b - Y_i = -j\omega\epsilon b\,\frac{t}{2}\,(1 - n^2)\,\frac{I_0^2(s\gamma b)}{I_0^2(\gamma b)} \\ \cdot\left(1 - \gamma^2 b^2 I_0^2(s\gamma b)\,\frac{t}{2}\,(1 - n^2)\left[\frac{K_0(s\gamma b)}{I_0(s\gamma b)} - \frac{K_0(\gamma b)}{I_0(\gamma b)}\right]\right)^{-1}. \end{aligned} \tag{17}$$

Equating this with eq. (16) at $n = 0$ yields the equation

$$\frac{1}{t} = \frac{1}{2}\,\theta^2 I_0^2(s\theta)\left[\frac{K_0(s\theta)}{I_0(s\theta)} + \frac{K_1(\theta)}{I_1(\theta)}\right], \tag{18}$$

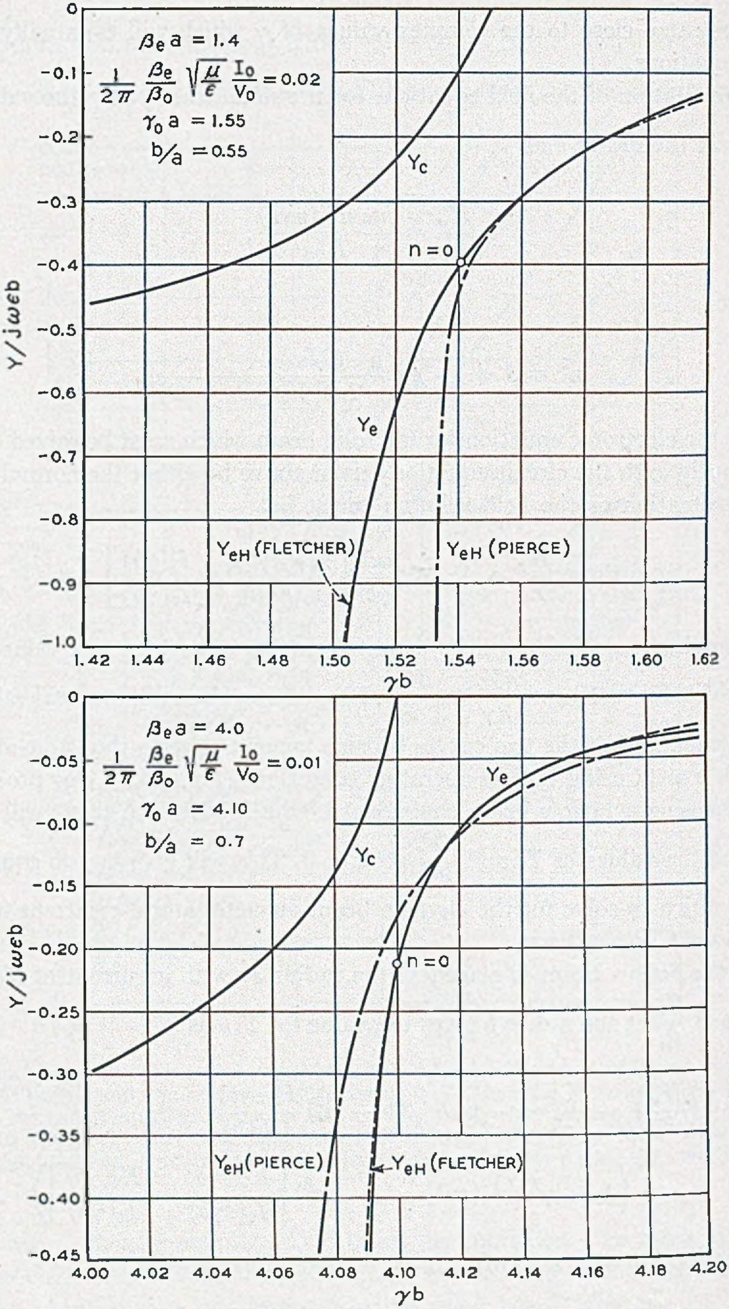


Fig. A6.2—Electronic admittance Y_e of a solid electron beam of radius b and circuit admittance Y_c of a helix of radius a plotted vs. real values of the propagation constant γ in the vicinity of where $\frac{dY_e}{d\gamma} = \frac{dY_c}{d\gamma}$ where complex solutions for γ are expected, for two typical sets of operating conditions. Plotted on the same graph is the electron admittance Y_{eH} for two equivalent hollow electron beams: the dashed curve (Fletcher) is matched to Y_e at $n = 0$, while the dot-dashed curve (Pierce, Appendix IV) is matched at $n = 1$ (off the graph).

where $\theta = \gamma_e b$ and γ_e is the value of γ at $n = 0$; i.e. for $\gamma_e \gg \beta_0$

$$\gamma_e = \beta_e + \sqrt{\frac{1}{\beta_0}\sqrt{\frac{\mu}{\epsilon}}\frac{\beta_e I_0}{2\pi b^2 V_0}} \approx \beta_e. \tag{19}$$

In the vicinity of $n = 0$, n varies very rapidly with γ, and hence matching $\left(\frac{\partial Y_e}{\partial n}\right)_\gamma$ is practically the same as matching $\frac{dY_e}{d\gamma}$. With this approximation eqs. (16) and (17) can be differentiated with respect to n and set equal at

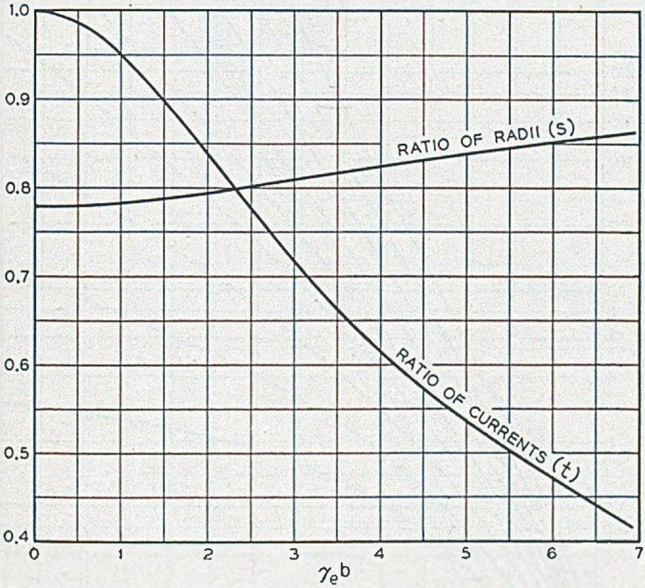


Fig. A6.3—Parameters of the hollow electron beam which is matched to the solid electron beam of radius b and current I_0 at $\gamma = \gamma_e \simeq \beta_e$, where $n = 0$. sb is the radius and tI_0 is the current of the equivalent hollow beam.

$n = 0$ to yield the second relation

$$\frac{1}{t} = \theta^2 I_0^2(\theta) I_0^2(s\theta)\left[\frac{K_0(s\theta)}{I_0(s\theta)} + \frac{K_1(\theta)}{I_1(\theta)}\right]^2 \tag{20}$$

Equations (18) and (20) can then be solved to give the implicit equation for s as

$$\frac{K_0(s\theta)}{I_0(s\theta)} = -\frac{K_1(\theta)}{I_1(\theta)} + \frac{1}{2I_1^2(\theta)} \tag{21}$$

and the simpler equation for t

$$t = \frac{4}{\theta^2}\frac{I_1^2(\theta)}{I_0^2(s\theta)}. \tag{22}$$

s and t are plotted as a function of θ in Fig. A6.3. The value of Y_{eH} using these values of s and t is compared in Fig. A6.2 with Y_e in the vicinity of where Y_e is almost tangent to Y_e for two typical sets of operating conditions.

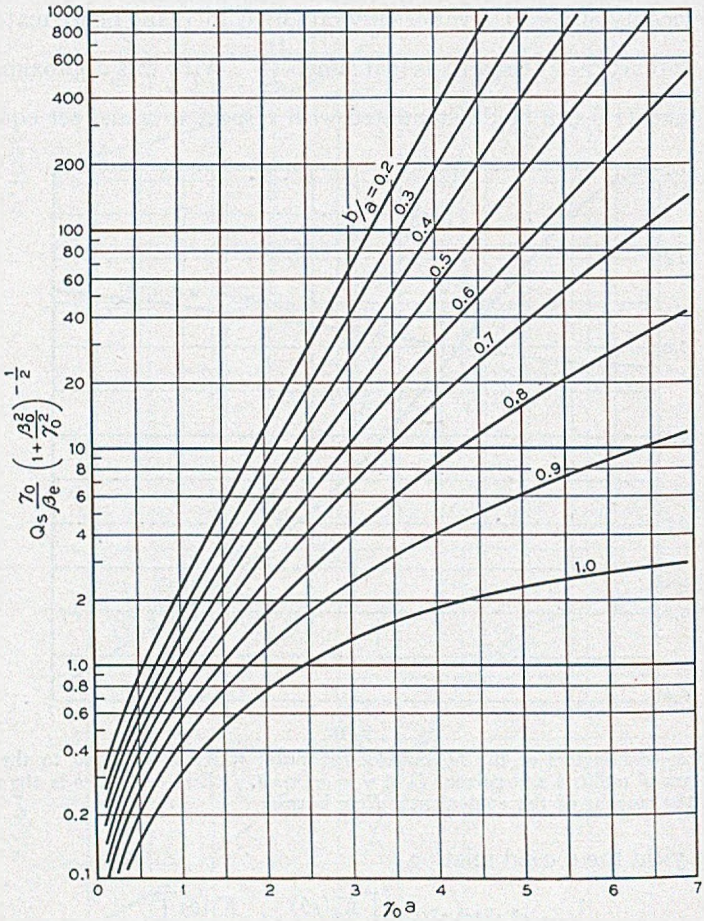


Fig. A6.4—Passive mode parameter Q_s for a solid beam of electrons of radius b inside a helix of radius a and natural propagation constant γ_0, obtained from the equivalent hollow beam parameters of Fig. 3 taken at $\gamma_e = \gamma_0$. All the normal mode solutions which have been found[(2), (3)] for a hollow beam will be approximately valid for a solid beam if Q is replaced by Q_s and K is replaced by K_s (Fig. 5).

It is of course possible to pick other criteria for determining an "equivalent" hollow beam. In Chapter XIV, in essence, Y_e and Y_{eH} were expanded in terms of $(1 - u^2)$ and the coefficients of the first two terms were equated. This has been done for the cylindrical beams, and the values of s and t found by this method determine values of Y_{eH} shown in Fig. A6.2. The greater departure from the true curve of Y_e would indicate that this approximation is not as good as that described above.

It is now possible to find the values of Q_s and K_s appropriate to the solid

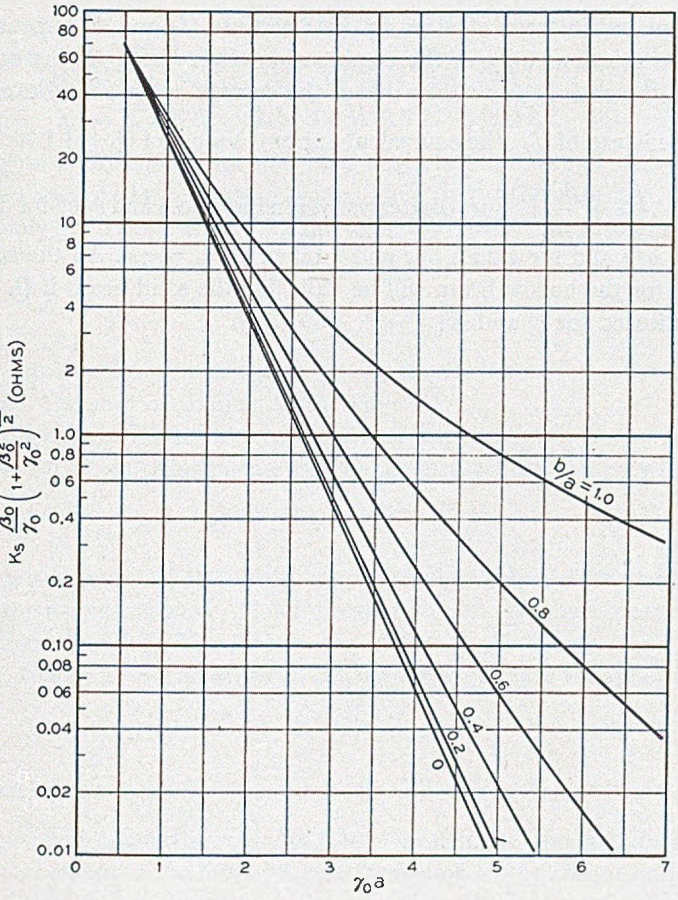


Fig. A6.5—Circuit impedance K_s for a solid beam of electrons of radius b inside a helix of radius a and natural propagation constant γ_0, obtained from the equivalent hollow beam parameters of Fig. 3 taken at $\gamma_e = \gamma_0$. K_s should replace $K = \dfrac{E^2}{2\beta^2_p}$ [(2), (3)] in order for the normal mode solutions for a hollow beam to be applicable to a solid beam.

beam. Thus if $Q\left(\gamma_0 a, \dfrac{b}{a}\right)$ and $K\left(\gamma_0 a, \dfrac{b}{a}\right)$ are the values for the hollow beam calculated from eqs. (9), (12) and (13),

$$Q_s = Q\left(\gamma_0 a, s\,\frac{b}{a}\right), \tag{23}$$

and

$$K_s = tK\left(\gamma_0 a,\ s\,\frac{b}{a}\right). \tag{24}$$

The t is placed in front of K in eq. (24) because tI_0 and K appear in the thin beam solutions only in the combination tI_0K. Using tK instead of K allows us to use I_0, the actual value of the current in the solid beam in the solutions instead of tI_0, the equivalent current. Values of $Q_s\,\frac{\gamma_0}{\beta_e}\left(1+\frac{\beta_0^2}{\gamma_0^2}\right)^{-1/2}$ and $K_s\,\frac{\beta_0}{\gamma_0}\cdot\left(1+\frac{\beta_0^2}{\gamma_0^2}\right)^{+3/2}$ are plotted vs. $\gamma_0 a$ in Figs. A6.4 and A6.5 for different values of b/a and for values of t and s taken at $\gamma_e = \gamma_0$. All the solutions obtained for the hollow beam will be valid for the solid beam if Q_s and K_s are substituted for Q and K.

APPENDIX VII

HOW TO CALCULATE THE GAIN OF A TRAVELING-WAVE TUBE

The gain calculation presented here neglects the effect at the output of all waves except the increasing wave. Thus, it can be expected to be accurate only for tubes with a considerable net gain. The gain is expressed in db as

$$G = A + BCN \tag{1}$$

Here A represents an initial loss in setting up the increasing wave and BCN represents the gain of the increasing wave.

We will modify (1) to take into account approximately the effect of the cold loss of L db in reducing the gain of the increasing wave by writing

$$G = A + [BCN - \alpha L] \tag{2}$$

Here α is the fraction of the cold loss which should be subtracted from the gain of the increasing wave. This expression should hold even for moderately non-uniform loss (see Fig. 9.5).

Thus, what we need to know to calculate the gain are the quantities

$$A, B, C, N, \alpha, L$$

A7.1 Cold Loss L db

The best way to get the cold loss L is to measure it. One must be sure that the loss measured is the loss of a wave traveling in the circuit and not loss at the input and output couplings.

A7.2 Length of Circuit in Wavelengths, N

We can arrive at this in several ways. The ratio of the speed of light c to the speed of an electron u_0 is

$$\frac{c}{u_0} = \frac{505}{\sqrt{V_0}} \tag{3}$$

where V_0 is the accelerating voltage. Thus, if ℓ is the length of the circuit and λ is the free-space wavelength and λ_g is the wavelength along the axis of the helix

$$\lambda_g = \lambda \frac{u_0}{c} \tag{4}$$

$$N = \frac{\ell}{\lambda_g} = \frac{\ell}{\gamma}\frac{c}{u_0} \tag{5}$$

Also, if $\mathfrak{L}_w$ is the total length of wire in the helix, approximately

$$N = \frac{L_w}{\lambda} \tag{6}$$

A7.3 The Gain Parameter C

The gain parameter can be expressed

$$C = \left(\frac{E^2}{\beta^2 P}\frac{I_0}{8V_0}\right)^{1/3} = \left(\frac{KI_0}{4V_0}\right)^{1/3} \tag{7}$$

Here K is the helix impedance properly defined. I_0 is the beam current in amperes and V_0 is the beam voltage.

A7.4 Helix Impedance K

In Fig. 5 of Appendix VI, $K\left(\frac{\beta_0}{\gamma_0}\right)\left(1 + \left(\frac{\beta_0}{\gamma_0}\right)^2\right)^{3/2}$ is plotted vs. $\gamma_0 a$ for values of b/a. K_s is the effective value of K for a solid beam of radius b, and a is the radius of the helix. γ_0 is to be identified with γ for present purposes, and is given by

$$\gamma_0 = \frac{2\pi}{\lambda_g}\left[1 - \left(\frac{\gamma_g}{\lambda}\right)^2\right]^{1/2} \tag{8}$$

where λ_g is given in terms of λ by (4). We see that in most cases (for voltages up to several thousand)

$$(\lambda_g/\lambda)^2 \ll 1 \tag{9}$$

and we may usually use as a valid approximation

$$\gamma_0 = \frac{2\pi}{\lambda_g} \tag{10}$$

and

$$\gamma_0 a = \frac{2\pi a}{\lambda_g} \tag{11}$$

As $\beta_0 = 2\pi/\lambda$, this approximation gives

$$1 + \left(\frac{\beta_0}{\gamma_0}\right)^2 = 1 + \left(\frac{\lambda_g}{\lambda}\right)^2$$

and we may assume

$$\left(1 + \left(\frac{\beta_0}{\gamma_0}\right)^2\right)^{3/2} = 1 \qquad (12)$$

Thus, we may take K_s as the ordinate of Fig. 5 multiplied by c/u_0, from (3), for instance.

The true impedance may be somewhat less than the impedance for a helically conducting sheet. If the ratio of the circuit impedance to that of a helically conducting sheet is known (see Sections 3 and 4.1 of Chapter III, and Fig. 3.13, for instance), the value of K_s from Fig. 5 can be multiplied by this ratio.

A7.5 The Space-Charge Parameter Q

The ordinate of Fig. 4 of Appendix VI shows $Q_s \dfrac{\gamma_0}{\beta_e}\left(1 + \left(\dfrac{\beta_0}{\gamma_0}\right)^2\right)^{-1/2}$ vs. γa for several values of b/a. Here Q_s is the effective value of Q for a solid beam of radius b. As before, for beam voltages of a few thousand or lower, we may take

$$\left(1 + \left(\frac{\beta_0}{\gamma_0}\right)^2\right)^{-1/2} = 1$$

The quantity β_e is just

$$\beta_e = \frac{2\pi a}{\lambda_g} \qquad (13)$$

and from (8) we see that for low beam voltages we can take

$$\beta_e = \gamma = \gamma_0$$

so that the ordinate in Fig. 4 can usually be taken as simply Q_s.

A7.6 The Increasing Wave Parameter B

In Fig. 8.10, B is plotted vs. QC. C can be obtained by means of Sections 3 and 4, and Q by means of Section 5. Hence we can obtain B.

A7.7 The Gain Reduction Parameter α

From (2) we see that we should subtract from the gain of the increasing wave in db α times the cold loss L in db. In Fig. 8.13 a quantity $\partial x_1/\partial d$, which we can identity as α, is plotted vs. QC.

A7.8 The Loss Parameter d

The loss parameter d can be expressed in terms of the cold loss, L in db, the length of the circuit in wavelengths, N, and C

$$d = \left(\frac{2.3L}{20}\right)\left(\frac{1}{2\pi NC}\right) \tag{14}$$

$$d = 0.0183 \frac{L}{NC} \tag{15}$$

A7.9 The Initial Loss A

The quantity A of (2) is plotted vs. d in Fig. 9.3. This plot assumes $QC = 0$, and may be somewhat in error. Perhaps Fig. 9.4 can be used in estimating a correction; it looks as if the initial loss should be less with $QC \neq 0$ even when $d \neq 0$. In any event, an error in A means only a few db, and is likely to make less error in the computed gain than does an error in B, for instance.

Technical Publications by Bell System Authors Other Than in the Bell System Technical Journal

*Progress in Coaxial Telephone and Television Systems.** L. G. Abraham.[1] *A.I.E.E., Trans.*, V. 67, pt. 2, pp. 1520–1527, 1948.

Abstract—This paper describes coaxial systems used in the Bell System to transmit telephone and television signals. Development of this system was started some time ago, with systems working before the war between New York and Philadelphia and later between Minneapolis, Minnesota and Stevens Point, Wisconsin. Various stages in the progress of this development have been described in previous papers and the telephone terminal equipment has been recently described. This paper will outline how the system works and discuss some transmission problems, leaving a complete technical description for a number of later papers.

Use of the Relay Digital Computer. E. G. Andrews and H. W. Bode.[1] *Elec. Engg.*, V. 69, pp. 158–163, Feb., 1950.

Abstract—This paper is concerned primarily with the operating features of the computer and its application to problems of scientific and engineering interest. The material herein has been derived largely from the experience gained with one of the computers during a trial period of about 5 months before final delivery. An effort was made during that time to try the machine out on a variety of difficult computing problems of varying character to obtain experience in its operation and to establish as well as possible what its range of usefulness might be.

Longitudinal Noise in Audio Circuits. H. W. Augustadt and W. F. Kannenberg.[1] *Audio Engg.*, V. 34, pp. 18–19, Feb., 1950.

Abstract—The words "longitudinal interference" have often been used to explain the origin of unknown noise in audio circuits with little actual regard to the source of the interference. In this respect, the usage of these words is similar to the popular usage of the word "gremlins". We attribute to gremlins troubles whose causes are unknown without much attempt to delve deeper into the matter. Similarly in the audio facilities field, many noise troubles are attributed to "longitudinal interference" or "longitudinals" or even simply "line noise" without a clear understanding of the nature of the trouble or the actual meaning of the terms. The noise trouble, however, still persists irrespective of the name applied to it until its causes are thoroughly understood and the correct remedial action is applied. This

* A reprint of this article may be obtained on request to the editor of the B.S.T.J.
[1] B.T.L.

paper describes and illustrates, with representative examples, various types of common noise induction in order to lead to an understanding of their nature. The paper includes, in addition, a discussion of simple remedies which may be employed for representative cases of noise troubles due to longitudinal induction.

Mobile Radio. A. Bailey.[3] *A.I.E.E., Trans.*, V. 67, pt. 2, pp. 923–931, 1948.

*Stabilized Permanent Magnets.** P. P. Cioffi.[1] *A.I.E.E., Trans.*, V. 67, pt. 2, pp. 1540–1543, 1948.

Abstract—Permanent magnets are stabilized against forces tending to demagnetize them, by partial demagnetization. It is shown that, after such stabilization, the magnet operates at a point on a secondary demagnetization curve. This curve may be treated identically as the major demagnetization curve is treated in ordinary magnet design problems. Formulas are developed for determining secondary demagnetization curves from the major demagnetization curve when stabilization is achieved by magnetization of the magnet before assembly, and by an applied magnetomotive force after magnetization in assembly.

It will be shown that, when the magnet is partially demagnetized for the purpose of stabilization, its operating point lies on a curve which, for convenience, will be called a secondary demagnetization curve. The object of this paper is to discuss the derivation of secondary demagnetization curves for given conditions of stability against demagnetizing forces and their applications to magnet design problems.

*Relay Preference Lockout Circuits in Telephone Switching.** A. E. Joel, Jr.[1] *A.I.E.E., Trans.*, V. 67, pt. 2, pp. 1720–1725, 1948.

Abstract—Occasions arise in telephone switching, particularly at common controlled stages, where calls compete for the use of equipment components or switching linkages. These call requests for service are received at random by circuits which must choose among and serve them on a one-at-a-time basis. Circuits which perform this function are known as "preference lockouts". Extensive use has been made of these circuits in manual, panel, and crossbar switching systems. This paper describes the design philosophies of relay preference lockout circuits based on some of these applications.

Piezoelectric Crystals and Their Application to Ultrasonics. W. P. Mason.[1] *Book*, New York, Van Nostrand, 508 pages, 1950.

*Television Terminals for Coaxial Systems.** L. W. Morrison, Jr.[1] *Elec. Engg.*, V. 69, pp. 109–115, February, 1950.

* A reprint of this article may be obtained on request to the editor of the B.S.T.J.
[1] B.T.L.
[3] A. T. & T.

Abstract—The broad features of operation of the L1 Coaxial System for the transmission of television have been discussed in a recent paper (L. G. Abraham, "Progress in Coaxial Telephone and Television Systems", AIEE Transactions, Vol. 67, pp. 1520–1527, 1948). It is the purpose of this paper to describe, in somewhat more detail, the factors influencing the design of the coaxial television terminals and the features of the equipment now in service in the Bell System's Television Network. The television terminals here described were placed into network service in 1947, but in basic form are similar to experimental models developed prior to the war and used in early television transmission studies over the coaxial cable..

Alternate to Lead Sheath for Telephone Cables. A. Paone.[3] *Corrosion*, V. 6, pp. 46–50, February, 1950.

*Bridge Erosion in Electrical Contacts and Its Prevention.** W. G. Pfann.[1] *A.I.E.E., Trans.*, V. 67, pt. 2, pp. 1528–1533, 1948.

Abstract—The size of the molten bridge which forms as two contacts separate depends upon the contact material and the current. The molten bridge has two diameters, one in each contact. By pairing dissimilar contact materials an asymmetric bridge is created, in which the bridge diameters are unequal and with which is associated a self-limiting transfer tendency. Under certain conditions the use of unlike pairs can prevent the continued transfer of material from one contact to the other.

*Chess-playing Machine.** C. E. Shannon.[1] *Sci. Am.*, V. 182, pp. 48–51, February, 1950.

*Military Teletypewriter Systems of World War II.** F. J. Singer.[1] Bibliography. *A.I.E.E., Trans.*, V. 67, pt. 2, pp. 1398–1408, 1948.

Abstract—This paper reviews the evolution of military teletypewriter communications since 1941 and briefly describes some of the important systems that were developed during the war by Bell Telephone System engineers for the armed forces.

*Optimum Coaxial Diameters.** P. H. Smith.[1] *Electronics*, V. 23, pp. 111–112, 114, February, 1950.

Abstract—The derivation of the optimum ratios is briefly described and optimum values are indicated to one part in ten thousand. In all cases the medium between conductors is assumed to be a gas with a dielectric constant approaching unity, and any effect of inner conductor supports upon the optimum conductor diameter ratio for a given property has been neglected.

*General Review of Linear Varying Parameter and Nonlinear Circuit Analysis.** W. R. Bennett.[1] *I.R.E., Proc.*, V. 38, pp. 259–263, March, 1950.

*A reprint of this article may be obtained on request to the editor of the B.S.T.J.
[1]B.T.L.
[3]A. T. & T.

Abstract—Variable and nonlinear systems are classified from the standpoint of their significance in communication problems. Methods of solution are reviewed and appropriate references are cited. The paper is a synopsis of a talk given at the Symposium on Network Theory of the 1949 National I.R.E. Convention.

Some Early Long Distance Lines in the Far West. W. Blackford, Sr.[4] and J. F. Hutton.[4] *Bell Tel. Mag.*, V. 28, pp. 227–237, Winter, 1949–50.

*Radio Propagation Variations at VHF and UHF.** K. Bullington.[1] *I.R.E., Proc.*, V. 38, pp. 27–32, January, 1950.

Abstract—The variations of received signal with location (shadow losses) and with time (fading) greatly affect both the usable service area and the required geographical separation between co-channel stations. An empirical method is given for estimating the magnitude of these variations at vhf and uhf. These data indicate that the required separation between co-channel stations is from 3 to 10 times the average radius of the usable coverage area, and depends on the type of service and on the degree of reliability required. The application of this method is illustrated by examples in the mobile radiotelephone field.

*Speaking Machine of Wolfgang von Kempelen.** H. Dudley[1] and T. H. Tarnoczy. *Acoustical Soc. Am., Jl.*, V. 22, pp. 151–166, March, 1950.

Perception of Speech and Its Relation to Telephony. H. Fletcher[1] and R. H. Galt.[1] *Acoustical Soc. Am., Jl.*, V. 22, pp. 89–151, March, 1950.

Abstract—This paper deals with the interpretation aspect and how it is affected when speech is transmitted through various kinds of telephone systems.

Vacuum Fusion Furnace for Analysis of Gases in Metals. W. G. Guldner[1] and A. L. Beach.[1] *Anal. Chem.*, V. 22, pp. 366–367, February, 1950.

Complex Stressing of Polyethylene. I. L. Hopkins,[1] W. O. Baker[1] and J. B. Howard.[1] *Jl. Applied Phys.*, V. 21, pp. 206–213, March, 1950.

Noise Considerations in Sound-Recording Transmission Systems. F. L. Hopper.[2] References. *S.M.P.E., Jl.*, V. 54, pp. 129–139, February, 1950.

*Radiation Characteristics of Conical Horn Antennas.** A. P. King.[1] *I.R.E., Proc.*, V. 38, pp. 249–251, March, 1950.

Abstract—This paper reports the measured radiation characteristics of conical horns employing waveguide excitation. The experimentally derived gains are in excellent agreement with the theoretical results (unpublished) obtained by Gray and Schelkunoff.

The gain and effective area is given for conical horns of arbitrary proportions and the radiation patterns are included for horns of optimum design.

* A reprint of this article may be obtained on request to the editor of the B.S.T.J.
[1] B.T.L.
[2] W. E. Co.
[4] Pac. T. & T.

All dimensional data have been normalized in terms of wavelength, and are presented in convenient nomographic form.

Microwaves and Sound. W. E. KOCK.[1] *Physics Today*, V. 3, pp. 20–25, March, 1950.

ABSTRACT—A recent development shows that obstacle arrays, modeled after the periodic structure of crystals, refract and focus not only electromagnetic waves, but sound waves as well. The behavior of periodic structures can be investigated by microwave and acoustic experiments on such models.

Interference Characteristics of Pulse-Time Modulation. E. R. KRETZMER.[1] *I.R.E., Proc.*, V. 38, pp. 252–255, March, 1950.

ABSTRACT—The interference characteristics of pulse-time modulation are analyzed mathematically and experimentally; particular forms examined are pulse-duration and pulse-position modulation. Both two-station and two-path interference are considered. Two-station interference is found to be characterized by virtually complete predominance of the stronger signal, and by noise of random character. Two-path interference, in the case of single-channel pulse-duration modulation, generally permits fairly good reception of speech and music signals.

*Electron Bombardment Conductivity in Diamond.** K. G. MCKAY.[1] *Phys. Rev.*, V. 77, pp. 816–825, March 15, 1950.

*Perception of Television Random Noise.** P. MERTZ.[1] References. *S.M.P.E., Jl.*, V. 54, pp. 8–34, January, 1950.

ABSTRACT—The perception of random noise in television has been clarified by studying its analogy to graininess in photography. In a television image the individual random noise grains are assumed analogous to photographic grains. Effective random noise power is obtained by cumulating and weighting actual noise powers over the video frequencies with a weighting function diminishing from unity toward increasing frequencies. These check reasonably well with preliminary experiments. The paper includes an analysis of the effect of changing the tone rendering and contrast of the television image.

*Loudness Patterns—A New Approach.** W. A. MUNSON[1] and M. B. GARDNER.[1] *Acoustical Soc. Am., Jl.*, V. 22, pp. 177–190, March, 1950.

Bell System Participation in the Work of the A.S.A. H. S. OSBORNE.[3] *Bell Tel. Mag.*, V. 28, pp. 181–190, Winter, 1949–50.

*New Electronic Telegraph Regenerative Repeater.** B. OSTENDORF, JR.[1] *Elec. Engg.*, V. 69, pp. 237–240, March, 1950.

Correlation of Gieger Counter and Hall Effect Measurements in Alloys Con-

* A reprint of this article may be obtained on request to the editor of the B.S.T.J.
[1] B.T.L.
[3] A. T. & T.

*taining Germanium and Radioactive Antimony 124.** G. L. PEARSON,[1] J. D. STRUTHERS,[1] and H. C. THEURER.[1] *Phys. Rev.*, V. 77, pp. 809–813, March 15, 1950.

*Optical Method for Measuring the Stress in Glass Bulbs.** W. T. READ.[1] *Applied Phys., Jl.*, V. 21, pp. 250–257, March, 1950.

Programming a Computer for Playing Chess. C. E. SHANNON.[1] References. *Phil. Mag.*, V. 41, pp. 256–275, March, 1950.

ABSTRACT—This paper is concerned with the problem of constructing a program for a modern electronic computer of the EDVAC type which will enable it to play chess. Although perhaps of no practical importance the question is of theoretical interest, and it is hoped that a satisfactory solution of this problem will act as a kind of wedge in attacking other problems of a similar nature and of greater significance.

Recent Developments in Communication Theory. C. E. SHANNON.[1] *Electronics*, V. 32, pp. 80–83, April, 1950.

ABSTRACT—In this paper the highlights of this recent work will be described with as little mathematics as possible. Since the subject is essentially a mathematical one, this necessitates a sacrifice of rigor; for more precise treatments the reader may consult the references.

A Symmetrical Notation for Numbers. C. E. SHANNON.[1] *Am. Math. Monthly*, V. 57, pp. 90–93, February, 1950.

*Capacity of a Pair of Insulated Wires.** W. H. WISE.[1] *Quart. Applied Math.*, V. 7, pp. 432–436, January, 1950.

*Echoes in Transmission at 450 Megacycles from Land-to-Car Radio Units.** W. R. YOUNG, JR.[1] and L. Y. LACY.[1] *I.R.E., Proc.*, V. 38, pp. 255–258, March, 1950.

*Simplified Derivation of Linear Least Square Smoothing and Prediction Theory.** H. W. BODE[1] and C. E. SHANNON.[1] *I.R.E., Proc.*, V. 38, pp. 417–425, April, 1950.

ABSTRACT—In this paper the chief results of smoothing theory will be developed by a new method which, while not as rigorous or general as the methods of Wiener and Kolmogoroff, has the advantage of greater simplicity, particularly for readers with a background of electric circuit theory. The mathematical steps in the present derivation have, for the most part, a direct physical interpretation, which enables one to see intuitively what the mathematics is doing.

*Helix Parameters Used in Traveling Wave-Tube Theory.** R. C. FLETCHER.[1] *I.R.E., Proc.*, V. 38, pp. 413–417, April, 1950.

ABSTRACT—Helix parameters used in the normal mode solution of the traveling-wave tube are evaluated by comparison with the field equations

* A reprint of this article may be obtained on request to the editor of the B.S.T.J.
[1] B.T.L.

for a thin electron beam. Corresponding parameters for a thick electron beam are found by finding a thin beam with approximately the same r-f admittance.

*Effect of Change of Scale on Sintering Phenomena.** C. Herring.[1] *Jl., Applied Phys.*, V. 21, pp. 301–303, April, 1950.

Abstract—It is shown that when certain plausible assumptions are fulfilled simple scaling laws govern the times required to produce, by sintering at a given temperature, geometrically similar changes in two or more systems of solid particles which are identical geometrically except for a difference of scale. It is suggested that experimental studies of the effect of such a change of scale may prove valuable in identifying the predominant mechanism responsible for sintering under any particular set of conditions, and may also help to decide certain fundamental questions in fields such as creep and crystal growth.

*Mode Conversion Losses in Transmission of Circular Electric Waves Through Slightly Non-Cylindrical Guides.** S. P. Morgan, Jr.[1] *Jl., Applied Phys.*, V. 21, pp. 329–338, April, 1950.

Abstract—A general expression is derived for the effective attenuation of circular electric (TE_{01}) waves owing to mode conversions in a section of wave guide whose shape deviates slightly in any specified manner from a perfect circular cylinder. Numerical results are in good agreement with experiment for the special case of transmission through an elliptically deformed section of pipe. The case of random distortions in a long wave guide line is analyzed and it is calculated, under certain simplifying assumptions, that mode conversions in a 4.732-inch copper pipe whose radius deviates by 1 mil rms from that of an average cylinder will increase the attenuation of the TE_{01} mode at 3.2 cm by an amount equal to 20% of the theoretical copper losses. The dependence on frequency of mode conversion losses in such a guide is discussed.

Acoustical Designing in Architecture. C. M. Harris[1] and V. O. Knudsen. *Book*, New York, John Wiley & Sons, Inc., 450 pages, 1950.

Abstract—This book is intended as a practical guide to good acoustical designing in architecture. It is written primarily for architects, students of architecture, and all others who wish a non-mathematical but comprehensive treatise on this subject. Useful design data have been presented in such a manner that the text can serve as a convenient handbook in the solution of most problems encountered in architectural acoustics.

* A reprint of this article may be obtained on request to the editor of the B.S.T.J.
[1] B.T.L.

Contributors to this Issue

R. V. L. Hartley, A.B., Utah, 1909; B.A., Oxford, 1912; B.Sc., 1913; Instructor in Physics, Nevada, 1909–10. Engineering Department, Bell Telephone Laboratories, 1913–50. Mr. Hartley took part in the early radio telephone experiments and was thereafter associated with research on telephony and telegraphy at voice and carrier frequencies. Later, as Research Consultant he was concerned with general circuit problems. Mr. Hartley is now retired from active service.

J. R. Pierce, B.S., in Electrical Engineering, California Institute of Technology, 1933; Ph.D., 1936. Bell Telephone Laboratories, 1936–. Dr. Pierce has been engaged in the study of vacuum tubes.

Claude E. Shannon, B.S., in Electrical Engineering, University of Michigan, 1936; S.M. in Electrical Engineering and Ph.D. in Mathematics, M.I.T., 1940. National Research Fellow, 1940. Bell Telephone Laboratories, 1941–. Dr. Shannon has been engaged in mathematical research principally in the use of Boolean Algebra in switching, the theory of communication, and cryptography.

George C. Southworth, B.S., Grove City College, 1914; Sc.D. (Hon.), 1931; Ph.D., Yale University, 1923. Assistant Physicist, Bureau of Standards, 1917–18; Instructor, Yale University, 1918–23. Editorial staff of The Bell System Technical Journal, American Telephone and Telegraph Company, 1923–24; Department of Development and Research, 1924–34; Research Department, Bell Telephone Laboratories, 1934–. Dr. Southworth's work in the Bell System has been concerned chiefly with the development of the waveguide as a practical medium of transmission. He is the author of numerous papers relating to a diversity of subjects such as ultra-short waves, short-wave radio propagation, earth currents, the transmission of microwaves along hollow metal pipes and dielectric wires and microwave radiation from the sun.