Ultra - High - Frequency Techniques

BY

J. G. BRAINERD (EDITOR)
Professor in the Moore School of Electrical Engineering, of the University of Pennsylvania

GLENN KOEHLER
Assistant Professor of Electrical Engineering University of Wisconsin

HERBERT J. REICH
Professor of Electrical Engineering University of Illinois

L. F. WOODRUFF
Associate Professor of Electrical Engineering Massachusetts Institute of Technology

SIXTH PRINTING

NEW YORK
D. VAN NOSTRAND COMPANY, INC.
250 FOURTH AVENUE

First Published, July, 1942
Fifth Printing, Corrected, September, 1942

Lithoprinted in U.S.A.
EDWARDS BROTHERS, INC.
ANN ARBOR, MICHIGAN
1942

PREFACE

This text is intended to cover the course outlined by representatives of some forty institutions who met at the Massachusetts Institute of Technology in November, 1941 to consider the demand, engendered by the war, for engineers and physicists with training in the ultra-high-frequency field. But despite the fact that the book is the result of war pressure, it covers no specific military material. Its contents will be useful in time of peace as well as war, and it represents essentially a reorientation of viewpoint rather than a choosing of topics for an excessively specialized purpose.

All the authors were members of the conference and in preparing the manuscript they have adhered closely to the syllabus produced there. The title of the book is the title chosen for the course, and is slightly ambiguous in that the text is more concerned with the bases of u-h-f techniques than with the techniques themselves, except in the laboratory manual of Ch. 15.

The book begins with a brief review of the elements of circuit theory and electron tubes. Thereafter, following the outline evolved by the conference, it presents in unified manner the material required as the minimum basis for technical work in the ultra-high-frequency field. In certain sections, subject matter is included which exceeds that required as the minimum basis, in order that the book may have even wider application. It also contains much necessary material from other fields, because u-h-f work requires the use of many low-frequency circuits and pieces of equipment. Consequently, considerable material from ordinary communication courses is presented (in condensed form), so that the student can work continuously through this book without referring to others as he proceeds.

The general "level" is that of senior students in electrical engineering and physics. Sufficient theoretical background is given for the various topics to make the presentations convincing, and to give the student an appreciation of at least some of the possibilities which each topic offers, the difficulties which may be encountered, and the actual or potential applications. Although a reasonable mathematical background is presented, particularly in those fields such as radiation and hollow wave guides where the senior student is likely to meet material entirely new to him, nevertheless most of the text can be read by a person seeking specific information without involvement in the more detailed developments of the theory. A brief laboratory manual is included covering the experiments which are an integral part of the course.

This text is designed as a strictly utilitarian tool for use in the standard techniques course. It has been prepared on short notice by men who have been well occupied with other duties, and the results of the quick preparation of the manuscript undoubtedly show in the text. The authors will welcome criticisms, not only to improve the present content of the book but also to make such changes in emphasis and content as will permit the text to be of greatest possible service at this time.

It has been convenient to make direct use of numerous existing illustrations, as well as indirect use of much published material. To the original producers of these the authors extend their thanks. In addition, Professor Reich has used sections from his "Principles of Electron Tubes" in Ch. 4 and a number of illustrations in Ch. 2. The authors are indebted to the McGraw-Hill Book Co., and in particular to Mr. James S. Thompson, Vice-President of that company, for permission to use this material.

Through an unfortunate mistake, credit was not given in the first printing to Dr. W. W. Hansen for original material furnished by him and used in Sec. 10-21. The authors regret this oversight.

This preface would not be complete without an acknowledgment of the work of Professor W. L. Barrow, who organized and presided over the conference from which the ultra-high-frequency techniques course grew, and from which arose the immediate cause for this book.

July, 1942 The Authors

It was not possible to have the manuscript read by each author, and such lack of coordination as may be found in the work is the fault of the editor. J.G.B.

EDITORIAL PREFACE

by

W. L. Barrow
Massachusetts Institute of Technology

Early in the summer of 1941 our country was in a period in which matters pertaining to its defense were receiving intensive consideration. We were then introduced to the variety of military and naval activity that has since become our daily fare. Many illuminating lessons were being learned from our allies as well as from our enemies, and our National Defense Program was being formulated into a coherent and clear picture. Although many radically new weapons, battle tactics, and military and naval devices were being continually presented to our view, no other single matter was more striking than the vital role of radio communication in the war. It became clear to many of us that the demand for trained specialists in electrodynamics and ultra-high-frequency techniques would shortly exceed the country's normal capacity to supply them. It was further evident that although the shortage of manpower would occur in both the upper and lower levels of competency, the shortage of men having a degree in engineering physics or its equivalent, as well as a specialized training appropriate to the new field, was a difficult problem to overcome. No substitute for the time element in preparing men of such qualifications can be devised.

A proposal for increasing the number of men trained at this high level was made by Dr. Karl T. Compton, Dean Edward L. Moreland, and others at the Massachusetts Institute of Technology in accordance with which a group of educational institutions throughout the country would provide uniform special courses for seniors in electrical engineering and physics. These courses, following adequate prerequisites, would provide a sufficient amount of the specialized knowledge necessary for immediate work in the field. The specific objective of this proposal was to create an immediate sourse of men trained at this high level for the following activities:

1. Army, Navy, and Marine Corps in their various branches.
2. Electronics Training Group Replacement, for a period of service in England.
3. Governmental research and development, such as Radiation Laboratory and Civil Service.
4. Industry, particularly in regard to its military and naval requirements.

5. Teaching, particularly to enlarge further the reserve of trained specialists.

further and most important objective was to organize a group of coöperating universities that could be utilized in ways dictated by the developments of the succeeding period. This proposal coincided with plans already in evolution in the Engineering Defense Training Bureau of the U.S. Office of Education, Federal Security Agency (renamed the Engineering, Science, Management Defense Training Bureau during the subsequent year). This Bureau under the leadership of Dean R. A. Seaton, gave its enthusiastic support and sponsorship to this program. Its foresight in anticipating the country's needs made possible the establishment and operation of this training program even prior to the eventful date of December 7 when the program of defense became a program of war.

A close coördination of the various institutions invited by the E.S.M.D.T. to participate in this program was necessary to insure the maintenance of certain minimum standards of instruction and of course content. This thought found expression in the brief course for instructors of electrodynamics that was held at the Massachusetts Institute of Technology at the request of E.S.M.D.T. A further reason for having instructors for these courses meet in conference was the newness and unavailability of much of the material on which the instruction was to be based. It was important that time be not wasted by the inclusion of inappropriate material and that the newer material, available only in scattered articles in periodicals, be made available in a format admitting of efficient instruction in both class and laboratory.

The instructors' conference was held at the Massachusetts Institute of Technology from October 27 through November 13, 1941, during which time numerous specialists gave talks on selected pertinent topics. Opportunities to visit laboratories were provided and demonstrations of equipment and technique in the newer fields were presented. The objectives of this conference were fourfold: the coördination of training, the accomplishment of a uniformity of instruction up to a certain minimum level, the dissemination of special technical information not available in textbooks, and the formulation of a syllabus of the course to be given in the forty institutions represented by delegates to the conference. This syllabus represents perhaps the most important result of the conference. Committees including all of the forty delegates assumed the responsibility of compiling the detailed technical syllabus for the course which was called Ultra-High-Frequency Techniques. The syllabus was carefully considered by the group as a whole and had their unanimous approval.

This conference provided the initial step toward the accomplishment of this training program. Notes from the talks of the various speakers were made available to aid instructors in their work and a further step was taken by consolidating at the Massachusetts Institute of Technology the purchases of apparatus for all approved institutions. It became evident to all delegates that a still further step would materially aid in the effective conduct of the program, namely; the preparation of a suitable text in which the pertinent topics were available in one volume and in a form adaptable to the needs of the institutions giving the course. Thus arose the conception of a book to be made available at the earliest possible moment. To accomplish this it was necessary to have the collaboration of a group of qualified instructors with actual experience in the course to prepare the written material and to find a publisher who would be willing to eliminate the traditional time consuming elements in book production. The first of these two requirements was met by associating together the four authors of this book, all of whom are pre-eminent in the field of electrical engineering and who are original workers and authorities in the particular phases of the subject which they undertook: Professor J. G. Brainerd of the University of Pennsylvania, Professor Glenn Koehler of the University of Wisconsin, Professor Herbert J. Reich of the University of Illinois, and Professor L. F. Woodruff of Massachusetts Institute of Technology. Professor Brainerd who has had wide experience in the preparation of text material assumed the additional responsibility of editing the material. This group has worked most intensively to complete the book in record time and their ready assumption of this extra work and effort should be viewed, in my opinion, as a personal and important contribution to the war effort.

The second requirement was satisfied by the D. Van Nostrand Company which provided freely of its secretarial and editorial staff, and by printing the book in offset rather than by the more laborious letterpress process, enabled it to be produced in record time. Thus the book is made available months earlier than would otherwise be possible.

It is hardly to be expected that such an intensive program would turn out a highly polished product and indulgence is asked from teachers and students alike who use the book. It is, however, our belief that the delay that would have been occasioned had the authors sought perfection would be inconsistent with the general objectives of the text and would not be justified under war conditions. In future editions such polish and modifications as are indicated by the experience of its use can be incorporated. The book, therefore, is presented without apologies in the sincere hope that it will be an appreciable factor in the training program, which has as its sole objective the defeat of the Axis.

July, 1942

INTRODUCTION

The reader will find on the whole that as he reads through this text he deals with elements rather than systems. The chapter titles for the most part emphasize this fact, and except in the chapters on transmitters and receivers, systems in which the elements are to be used do not appear. Thus it becomes necessary to keep in mind the very great number of possible uses of the various topics as they arise.

In general, a system may be said to transform an original signal or stimulus into a useful result at a specified point or points. The stimulus may be of acoustic, electric, thermal, light, mechanical, or other origin. It in some way enters the system, is changed to an electrical representative which is modulated, amplified, limited, cut, transmitted, detected, and otherwise purposely changed, and ultimately emerges in useful form.

The pickup devices--microphones, photoelectric cells, thermocouples, and all the numerous other means for changing a stimulus of non-electric kind into an electric signal--are not treated here. Neither are the non-electrical devices such as loudspeakers, curve-tracers, etc., which are the ultimate receivers yielding the useful result. However, many electrical circuits and devices of importance as pickups or reproducers in modern radio applications are included.

It is highly desirable to become accustomed to thinking in terms of block diagrams. A block diagram of a system is one in which each essential major element is indicated by a large rectangle, and labeled appropriately. Several block diagrams are shown in Ch. 9, and others can be constructed to suit the problem at hand. For example, a long-distance telephone conversation in which carrier telephony is used would show blocks of the pickup (caller's telephone), local central office, long-distance office, local oscillator, modulator, transmission line with repeaters (amplifiers), long-distance office at the receiving point, filters, detector, amplifier, local office at the receiving point, and ultimate receiver (telephone of the person called). If one also considers television, picture transmission, telegraph, radio broadcasting, marine and aeronautical beacons, blind-landing systems, and so on, it is seen that block diagrams may grow almost ad lib. But, frequently, block diagrams may not cover such broad topics. A simple instrument--for example, a recording strain gauge of electrical design--may be outlined in a block diagram and thus patterned to a plan of units, many of which may possibly be constructed independently of the others.

It may be asked, "Why are high frequencies used?" There are numerous reasons, some of the more important of which are the need of high-frequency currents for radiation, the necessity of spreading over a large frequency spectrum to provide many channels, and the size of equipment in relation to wavelength.

With this brief introduction the reader will be left to a perusal of the discussions of the many elements or components which can possibly be combined to obtain a desired result in a given case.

It is desirable, however, to mention several usages usually frowned upon which have been accepted here. One is the use of kc and Mc or kilocycles and megacycles when kc/s and Mc/s or kilocycles per second and megacycles per second are meant. Until some simpler name than cycles per second is adopted in the English-speaking countries, it is inevitable that kilocycles or kc, and megacycles or Mc, will be used in oral transmission. The popular custom has been followed here. Another designation, not often approved but found in this book, is that of using the words frequencies, high frequencies, and ultra-high frequencies instead of the words current or voltage components of high frequency. For example, the statement "the high frequencies pass through an impedance" is employed many times to mean "the high-frequency components of the current pass through an impedance." Likewise, the term wave has at times been used intentionally in a generic sense, meaning either voltage or current or both. And finally, direct-current has been used as an adjective in a general sense, leading to such expressions as d-c voltage.

CONTENTS

Chapter I

LINEAR CIRCUIT ANALYSIS

Ultra-high-frequency circuits and devices require in their design, operation, and maintenance techniques which differ in many respects from those which serve at the lower frequencies. Yet the same basic principles apply at the ultra-high frequencies, and in addition, devices which depend in part on currents of ultra-high frequency may have in them elements or sections which operate at much lower frequencies, or even with direct current. Thus efficient work in the u-h-f field depends not on a knowledge of the ultra-high frequency region alone, but on the theory and techniques in virtually all the frequency ranges up to the u-h-f as well.

It is expected that the student will have some acquaintance with the methods of linear circuit analysis and with thermionic devices and their applications. Accordingly, only a very compact presentation of underlying principles will be given here, arranged as much for purposes of easy reference as for study material.

1-1. Alternating Currents.- Alternating currents and voltages in circuits having fixed parameters of resistance, inductance, and capacitance will be the subject of our initial investigation. The ideal alternating wave has a sinusoidal variation with time; for example

$$e = E \sin (\omega t + \varphi) \text{ instantaneous volts} \qquad (1\text{-}1)$$

as expressed in terms of a peak or maximum voltage E, an angular velocity ω equal to 2π times frequency, and a phase angle φ. An alternating or cyclical wave may have some other shape than a pure sinusoid, but so long as it is periodic it may be resolved into its sinusoidal or harmonic components, each of which will have the form of eq. (1-1), but with its individual values of E, ω and φ. If we are dealing with linear circuits, each sinusoidal component applied to the circuit will produce its response independent of any others, and therefore a linear circuit theory based on sinusoidal wave forms will serve also when the wave form is not sinusoidal.

1-2. Complex Quantities and Exponentials.- The various operations which must be performed on expressions such as eq. (1-1), if applied directly, lead to rather cumbersome expressions, and so it is common usage to represent the entire equation by a rotating vector E as shown in Fig. 1-1. The vector E at the left, with length equal to the maximum value of the sine wave, is shown in a position such that its projection on

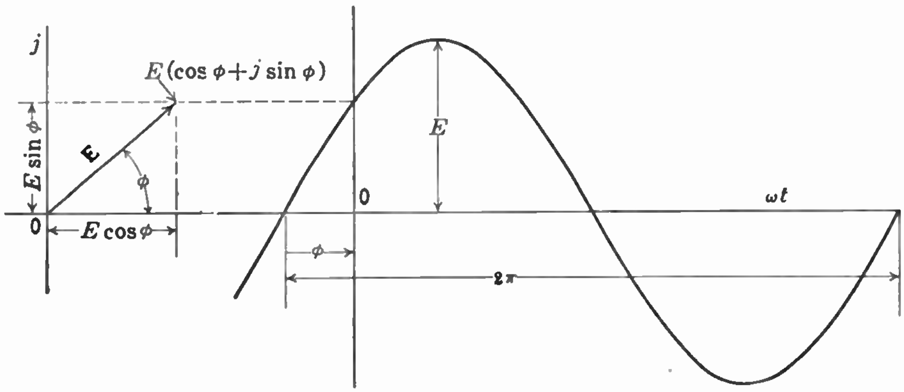


FIG. 1-1. Sinusoidal voltage wave and representation by a vector and by a complex number.

the vertical axis is equal to the instantaneous value of the sine wave when t = 0. If the vector rotates counterclockwise with angular velocity ω, the projection on the vertical at any time t is equal to E sin (ωt + φ) and so to the ordinate of the sine wave. In this sense, the rotating vector represents the entire sine wave, and two or more sine waves of the same frequency may be combined by addition or subtraction of their representative vectors.

Since rotating vectors are commonly drawn with their starting points at the origin, the terminal point only needs to be specified in order to specify the vector and the original sine wave. The terminal point is conveniently specified in terms of a single complex number a + jb, where $j \equiv \sqrt{-1}$, with the real portion a understood to specify the coordinate of the point along the real axis, usually the horizontal; and the size b of the imaginary portion specifying the other coordinate. The addition and subtraction of complex numbers follow the same laws as those for combining vectors in a plane.

Instead of considering in turn the rotating vector as the analog of the sine wave, and the complex number as the analog of the vector, the same result is obtained from the identity.[1]

$$e^{j(\omega t + \varphi)} = \cos (\omega t + \varphi) + j \sin (\omega t + \varphi) \qquad (1\text{-}2)$$

1. The e in (1-2,3,4 etc.) is the base of the natural logarithms. This is standard notation, but leads to a conflict with the use of e for voltage.

so that the original sine function is the size of the imaginary portion of the exponential. If the original wave had been a cosine function, it would have been the real part of the exponential. The advantage of the exponential form over the sine or cosine form manifests itself when derivatives or integrals are involved in the circuit; and is due to the fact that taking derivatives and integrals of exponentials does not alter the form of the original function. Actually, the sine and cosine functions may be expressed in exact exponential form by the identities

$$\sin \omega t = \frac{e^{j\omega t} - e^{-j\omega t}}{2j}; \tag{1-3}$$

$$\cos \omega t = \frac{e^{j\omega t} + e^{-j\omega t}}{2}; \tag{1-4}$$

but the writing of a great many characters is saved by writing only one of the exponentials, or none at all. Thus we have a variety of possible expressions for a wave, such as

$$\begin{aligned} e &= E \cos(\omega t + \varphi) \\ &= E \sin(\omega t + \varphi + \pi/2) \\ &= E(\cos\varphi \cos\omega t - \sin\varphi \sin\omega t) \qquad (1\text{-}5) \\ &= E\,\frac{e^{j(\omega t+\varphi)} + e^{-j(\omega t+\varphi)}}{2} \\ &= \text{Real part of } E\, e^{j(\omega t+\varphi)}. \end{aligned}$$

The five foregoing expressions state the value of the instantaneous voltage e as explicit functions of time. The words "Real part of" in the last expression of (1-5) are sometimes abbreviated to "Re," but more often omitted altogether, although understood, leaving only $E\, e^{j(\omega t+\varphi)}$.

The expression for the rotating coplanar vector $\underline{E}$ is variously written as

$$\begin{aligned} \underline{E} &= E\, e^{j\varphi}. \\ &= E(\cos\varphi + j\sin\varphi) \\ &= E\,\underline{/\varphi} \end{aligned}$$

Throughout this text a complex number or vector will be represented by an underscored symbol; the same symbol without the underscoring will represent the size only of the quantity. The

---

Footnote continued

The context will usually enable the two uses to be easily distinguished; for example in the last of the group of equations denoted (1-5) the two e's appear but the meaning of each is evident.

vector $\underline{E}$ is the complex equivalent of the instantaneous e in the manner already explained, and illustrated in Fig. 1-1, but there is not a true mathematical equality between e and $\underline{E}$.

1-3. Impedances and Admittances.- There are just three basic forms of circuit parameters which tend to limit current flow. These are resistance (R), inductance (L), and capacitance (C). If a current i is flowing through a series connection of R, L, and C, the voltage drop e is equal to

$$e = Ri + L\frac{di}{dt} + \frac{1}{C}\int i\,dt \quad \text{volts} \tag{1-6}$$

R being expressed in ohms, L in henrys and C in farads. If $i = Ie^{j\omega t}$, then by substitution

$$e = \left(R + j\omega L + \frac{1}{j\omega C}\right) Ie^{j\omega t} \tag{1-7}$$

and the quantity in parenthesis is called the impedance of the connection, and designated by the symbol $\underline{Z} = R + jX$, X being called the reactance and equal to $\omega L - \frac{1}{\omega C}$. It will be noted that $\underline{Z}$ is in general a complex number, and the ratio of the size of $\underline{E}$ to the size of $\underline{I}$ is the size of $\underline{Z}$. Complex numbers, like vectors, are indicated by underscored symbols; the size of the number by the same symbol without underscoring. The angle of $\underline{Z}$ is the angle between the voltage vector $\underline{E}$ and the current vector $\underline{I}$.

Inverting the expression

$$\underline{Z} = \frac{\underline{E}}{\underline{I}} = R + jX \quad \text{complex ohms} \tag{1-8}$$

we may write the admittance

$$\underline{Y} = \frac{1}{\underline{Z}} = \frac{\underline{I}}{\underline{E}} = G + jB \quad \text{complex mhos,} \tag{1-9}$$

where G is the conductance and B the susceptance. In terms of the series resistance R and reactance X of the same branch,

$$G = \frac{R}{Z^2} \qquad \text{mhos;} \tag{1-10}$$

$$B = \frac{-X}{Z^2} \qquad \text{mhos.} \tag{1-11}$$

A series connection of several impedances $\underline{Z}_1$, $\underline{Z}_2$, ..., $\underline{Z}_n$ offers a total impedance of $\underline{Z}_1 + \underline{Z}_2 + \ldots + \underline{Z}_n$ complex ohms; and a parallel connection of several admittances $\underline{Y}_1$, $\underline{Y}_2$, ... $\underline{Y}_n$ offers a total admittance of $\underline{Y}_1 + \underline{Y}_2 + \ldots + \underline{Y}_n$ complex mhos. For several impedances $\underline{Z}_1$, $\underline{Z}_2$, ... $\underline{Z}_n$ in parallel, the equivalent impedance $\underline{Z}$ is

$$\underline{Z} = \frac{1}{\frac{1}{\underline{Z}_1} + \frac{1}{\underline{Z}_2} + \ldots + \frac{1}{\underline{Z}_n}} \quad \text{complex ohms.} \tag{1-12}$$

For only two branches, eq. (1-12) becomes

$$\underline{Z} = \frac{\underline{Z}_1\underline{Z}_2}{\underline{Z}_1 + \underline{Z}_2} \tag{1-13}$$

1-4. Kirchhoff's Laws (Loop and Node Equations).- When we have to deal with a network of impedances, the current and voltage relations must be solved by the use of the two Kirchhoff laws of zero voltage around any loop, and zero net current to any junction or node, combined with the impedance relation of voltage to current in each branch. The forms of the basic equations on which network solutions must depend are

$$\Sigma\underline{I} = 0 \text{ at any junction (node);} \tag{1-14}$$

$$\Sigma\underline{E} = 0 \text{ around any loop.} \tag{1-15}$$

The $\underline{E}$'s must include not only the $\underline{Z}\ \underline{I}$ drops but also any generated or other voltage of the same frequency.

The $\underline{I}$'s may be the actual currents in each branch, or optionally the equations may be written in terms of hypothetical "loop" currents which after solution are combined by vector addition or subtraction in such branches as are common to two or more loops.

1-5. Duality of Series and Parallel Circuits.- The mutual relations between complex voltage and current, in the three basic types of circuit parameter, may be written in a form to bring out the reciprocal nature of the relations which exist.

$$\left.\begin{aligned} \underline{E} &= R\,\underline{I} \\ \underline{E} &= j\omega L\underline{I} \\ \underline{E} &= \underline{I}/j\omega C \end{aligned}\right\} \qquad \Sigma\underline{E} = 0 \text{ around a loop} \tag{1-16}$$

$$\left.\begin{aligned} \underline{I} &= G\,\underline{E} \\ \underline{I} &= j\omega C\underline{E} \\ \underline{I} &= \underline{E}/j\omega L \end{aligned}\right\} \qquad \Sigma\underline{I} = 0 \text{ at a junction} \tag{1-17}$$

The second group of equations (1-17) is obtained from the first group (1-16) by substituting $\underline{I}$ for $\underline{E}$, $\underline{E}$ for $\underline{I}$, G for R, C for L and L for C. In all cases it is possible, starting with a circuit to which a group of equations such as (1-16) is appropriate, to determine a dual of that circuit to which the appropriate equations will be those obtained by the substitutions just enumerated. It is to be observed that the dual of a series branch containing resistance and inductance is a parallel connection of conductance and capacitance. The dual of a series R, L, C branch is a parallel combination of G, C, and L, actually three individual branches. Note that the circuit shown in Fig. 1-9 is not the dual of a simple series R, L, C circuit. In other words, the circuit usually called a parallel resonant circuit is not the dual of a series resonant circuit except in the special case in which resistances can be neglected. The dual of a loop consisting of several series branches is a junction having radiating from it the duals of these branches. The dual of voltage is current; and of a constant-voltage source, a constant-current source.

In so far as calculations are concerned, nothing is gained by the analysis of the dual of a circuit instead of the original, since the equations are identical, but there may be already available a solution of a circuit which is the dual of one to be analyzed. In such a case, need for a second analysis is eliminated. For certain purposes of measurement or control, when a specified function must be realized as a relation between voltage and current, there may be an option whether to use a specific circuit or its dual. Choice can then be based entirely on considerations of cost and convenience.

1-6. Principle of Superposition.- In a linear system, each applied force produces a response independent of the response due to any other applied force, and the total response is the sum of the responses due to all the applied forces.

If a linear[2] network has a response characteristic such that a voltage $\underline{E}_a$ applied in series with any branch a causes a flow of current $\underline{I}_{ca}$ in branch c, and if also a voltage $\underline{E}_b$ applied in series with a branch b causes a flow of current $\underline{I}_{cb}$ in branch c, then when both $\underline{E}_a$ and $\underline{E}_b$ are applied simultaneously they cause a total current $\underline{I}_{ca} + \underline{I}_{cb}$ in branch c.

A useful application of this relation is in the calculation of the current which will flow in a new branch of impedance $\underline{Z}_{ab}$ to be connected between two points a and b of a network having any number of voltage sources. If a hypothetical generator, of voltage equal and opposite to that existing at the connection point, is assumed to be in series with the new branch ab, then no current would flow. This zero current may be regarded as the sum of the current which would have flowed in the absence of the hypothetical generator, and the current produced by the hypothetical generator in the new branch and the original

2. Most circuits containing only lumped parameters, transmission lines, some tube circuits, etc. are linear; circuits containing iron-core reactors, many tube circuits, etc. are non-linear.

network in the absence of any other generated voltages. The current in the new branch may be found then as the negative of that caused by the single hypothetical generator and is the quotient of the voltage of the system between the points of connection a and b before the connection was made, divided by the sum of the new branch impedance $\underline{Z}_{ab}$ plus the impedance looking into the original network from the point of connection. This application of the superposition principle is sometimes called Thevenin's theorem.

A similar artifice may be used in connection with the calculation of current which will flow in an existing branch A when there is inserted in series with that branch a new impedance $\underline{Z}$. It may be shown that the current in the branch after the insertion of the impedance $\underline{Z}$ is equal to $\underline{I}_o/(1 + \underline{Z}\ \underline{Y}_{AA})$, where $\underline{I}_o$ is the current before the insertion, and $\underline{Y}_{AA}$ the admittance looking into the network from A, before the addition of $\underline{Z}$. This is sometimes called Norton's theorem.

The superposition principle is often useful in determining the response as a function of time, due to any arbitrary force function of time. In order to apply it for such a purpose, it is necessary to know the response to some simple basic force function, such as the unit step function. This is a function which has a value of zero before and unity after the time t = 0. This use involves transient phenomena, and will be discussed later in the chapter.

1-7. Reciprocity Theorem.- The reciprocity theorem, due to Rayleigh, states that in any linear network there will be the same response $\underline{I}$ in a branch B due to a force $\underline{E}$ applied in any branch A, as will exist when the positions are reversed, that is, when the force $\underline{E}$ is applied in branch B and the response measured in branch A.

1-8. Equivalent Constant-Voltage and Constant-Current Generators.- A generator or linear amplifier which develops a fixed internal voltage $\underline{E}$ and has an internal impedance $\underline{Z}$ is equivalent in its external characteristics to a constant-current generator developing a fixed current $\underline{E}/\underline{Z}$, and shunted by the same impedance $\underline{Z}$. See Fig. 1-2.

The constant-current equivalent circuit is often much simpler to analyze, because with it all the various loads on the generator may often be reduced to a number of simple parallel branches; otherwise series-parallel circuits would have to be handled.

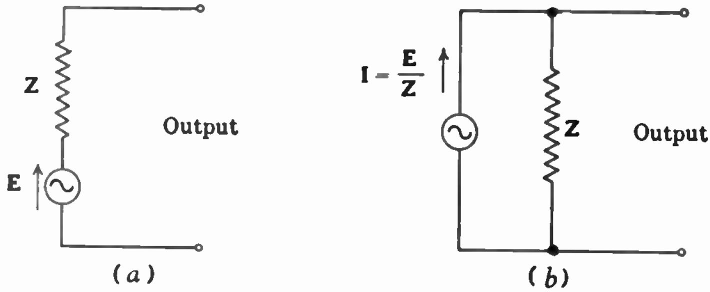


FIG. 1-2. Equivalent generator circuits.
(a) Constant E with series Z;
(b) Constant I with shunt Z.

1-9. Maximum Power Transfer.- In most communication circuits operating at low and medium power levels it is usually more important to obtain maximum power output from a generator or source of any kind, than to operate with a very high efficiency or ratio of output to input powers. By Thevenin's Theorem a source is nearly always representable by a simple circuit such as that of Fig. 1-2a, with fixed internal impedance and fixed voltage developed.

Suppose first that the internal impedance is representable as a simple ohmic resistance R; for example, the plate resistance of a vacuum-tube amplifier. If the output supplies a load of resistance R_L, then obviously the power P transferred to the load will be $R_L I^2$, or, in terms of E, R and R_L,

$$P = \left(\frac{E}{R + R_L}\right)^2 R_L. \qquad (1\text{-}18)$$

To find the condition for maximum transfer of power P we differentiate P with respect to R_L and equate the derivative to zero, whence we find:

$$R_L = R; \qquad (1\text{-}19)$$

$$P_{max} = \frac{E^2}{4R} \text{ watts}. \qquad (1\text{-}20)$$

If the internal impedance $\underline{Z}$ is complex, equal to R + jX, then obviously the load should have a reactance component of -X, if a maximum power transfer is to be attained, and

$$\underline{Z}_L = R - jX; \tag{1-21}$$

$$P_{max} = \frac{E^2}{4R} \text{ watts.} \tag{1-22}$$

If there is a constraint on the load impedance $\underline{Z}_L$ so that it must be of some arbitrary specified power-factor angle θ, in general different from that of $\underline{Z}$ whose angle will be called φ, then the power transferred is

$$P = \frac{E^2 Z_L \cos\theta}{(Z\cos\varphi + Z_L\cos\theta)^2 + (Z\sin\varphi + Z_L\sin\theta)^2} \text{ watts,} \tag{1-23}$$

which has a maximum for

$$Z_L = Z, \tag{1-24}$$

and

$$P_{max} = \frac{E^2\cos\theta}{2Z[1 + \cos(\varphi - \theta)]} \text{ watts.} \tag{1-25}$$

Equation (1-25) is a more general form from which (1-20) may be checked by setting $\theta = 0$ and $\varphi = 0$; and (1-22) also may be checked by setting $\varphi = -\theta$.

1-10. Frequency Characteristics of Simple Networks.- Non-dissipative networks are made up of elements of inductance and capacitance, the reactances of which are respectively ωL and $-1/\omega C$. It may be observed that the derivative of each of these reactances with respect to ω is positive. This holds in general for any linear network, and for any reactance associated with the network, whether it be a self or a transfer[3] reactance. Hence a plot of reactance as ordinates against ω as abscissas will always produce curves which have a positive or upward slope throughout the entire range $0 < \omega < \infty$. The magnitude of the derivative of each of the basic reactance terms ωL and $-1/\omega C$, taken with respect to ω, is equal to X/ω, and this value of derivative is the minimum to be found in any linear network.

Since the slope of the reactance X, which may be written

3. Transfer impedance is the factor relating the current at one point in a network to a voltage at some other place in the network; self-impedance relates the current through to the voltage across a particular branch.

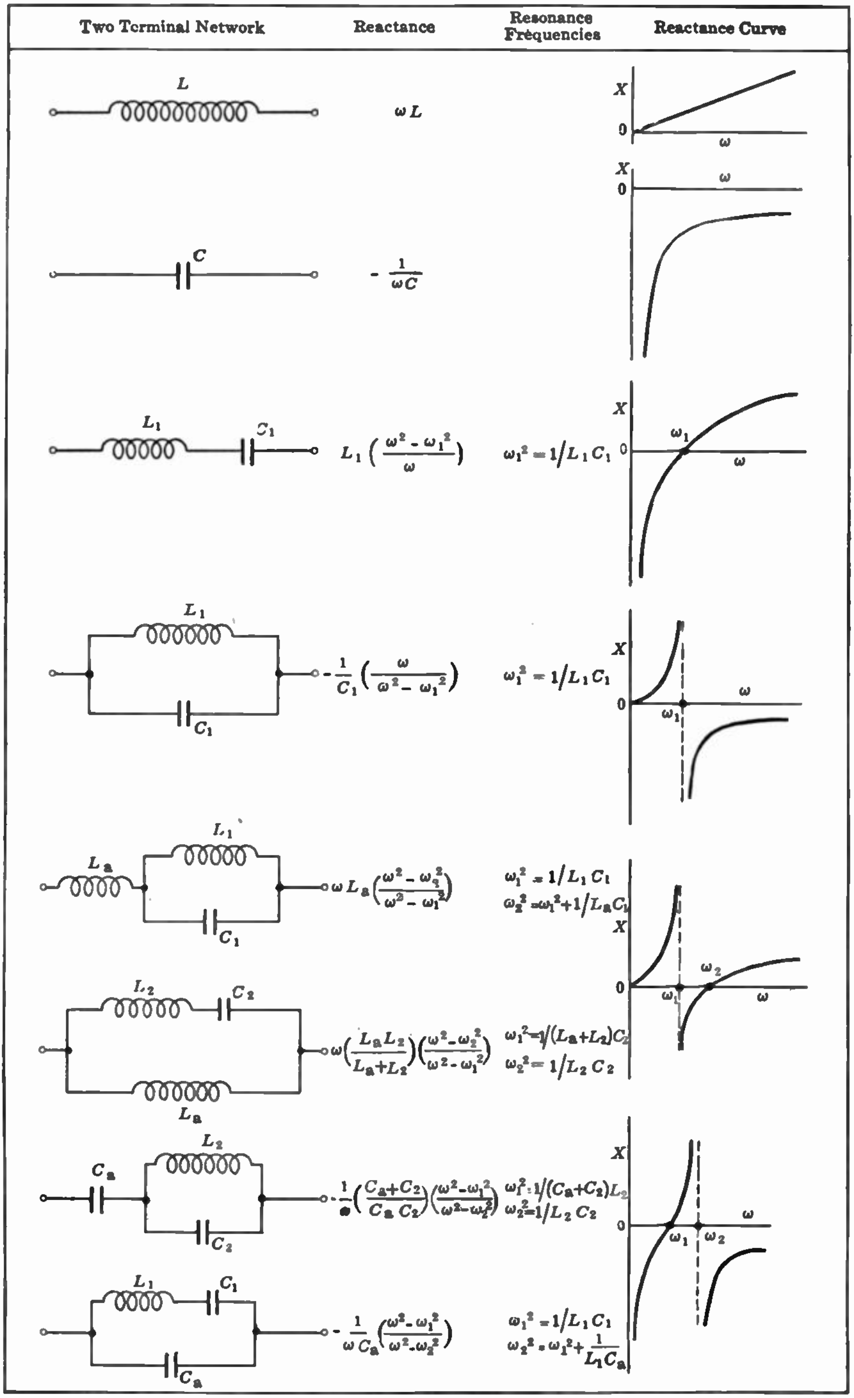
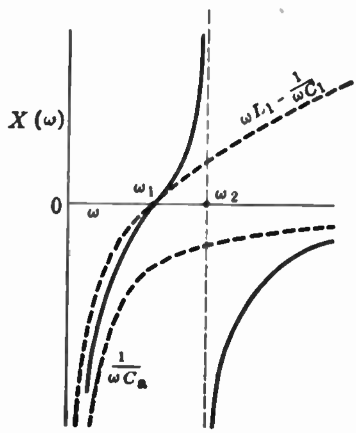


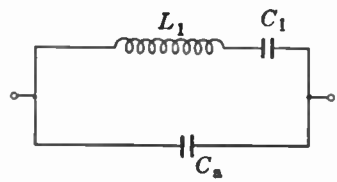
FIG. 1-3. Frequency characteristics of basic circuits

$X(\omega)$ to show that X depends on ω, is always positive, it follows that zeros of the function will be separated by infinite values, where the $X(\omega)$ goes on up to infinity and begins again at minus infinity.

Frequency characteristics of some simple non-dissipative networks are shown in Fig. 1-3.

The forms of the graphs of the reactance functions of simple combinations of the basic circuits may be deduced readily by combining the graphs of the first four elements shown in Fig. 1-3. For example, in the circuit of Fig. 1-4 there are two





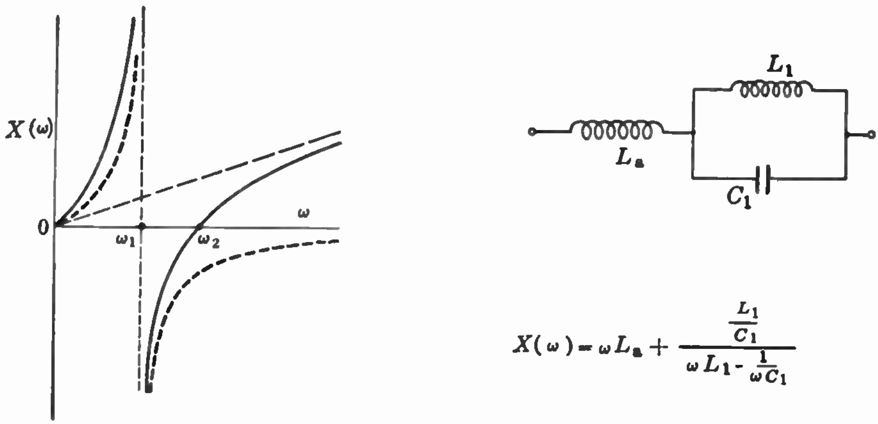
$$X(\omega) = \frac{-\frac{1}{\omega C_a}\left(\omega L_1 - \frac{1}{\omega C_1}\right)}{\omega L_1 - \frac{1}{\omega C_1} - \frac{1}{\omega C_a}}$$

$$= -\frac{1}{\omega C_a}\left(\frac{\omega^2 - \omega_1^2}{\omega^2 - \omega_2^2}\right)$$

FIG. 1-4. Frequency characteristic of a simple non-dissipative network.

parallel branches having characteristics shown in the previous figure, and repeated here as dotted curves. When the reactances of the two branches are equal in size and of opposite sign, the combined reactance will be infinite, as shown on the graph at ω_2. When the reactance of the upper branch is zero, at ω_1, the combined reactance is zero. For $\omega < \omega_1$, each branch is capacitive (reactance negative); hence the combination resembles at any one frequency two capacitors in parallel; for $\omega > \omega_2$ the combination resembles at any one frequency an inductor and a capacitor in parallel.

Again in the series-parallel circuit of Fig. 1-5, the graphs of the $X(\omega)$ functions of the parallel connection and of the series connection are shown separately by dotted lines. The full lines, whose ordinates are the simple sums of the corresponding ordinates of the dotted lines, comprise the graph of the reactance of the entire circuit.



$$X(\omega) = \omega L_a + \frac{\frac{L_1}{C_1}}{\omega L_1 - \frac{1}{\omega C_1}}$$

$$= \omega L_a \left(\frac{\omega^2 - \omega_2^2}{\omega^2 - \omega_1^2} \right)$$

FIG. 1-5. Frequency characteristic of a non-dissipative two-terminal network.

It is possible to graph in one plane the impedance or reactance characteristics of any network which has no resistance elements, but since this ideal non-dissipative condition is physically unrealizable, it must be recognized that actual impedance functions $\underline{Z}(\omega)$ are complex functions, and in fact never have a zero or an infinity provided there is some resistance (and there always is) in each branch. When the resistances are small, they have usually no material effect on the impedance except near the points where $X(\omega)$ is zero or infinity in the corresponding non-dissipative circuit. The function $\underline{Z}(\omega)$ has a size $Z(\omega)$ which is approximately given by inverting all the negative sections of the $X(\omega)$ functions for the dissipationless circuit, and altering the zeros and infinities by rounding off the sharp junctions, if any exist, by small cusps, thus eliminating zeros and infinities.

The reactance function $X(\omega)$ in passing through the zero axis, and changing sign, could be described also as undergoing a phase shift of π radians or 180 degrees. The same change occurs at the infinities; that is, another change of 180 degrees. In the general impedance function, a plot of the angle of $\underline{Z}$ as ordinate against ω as abscissa will exhibit a very steep slope where the size Z of $\underline{Z}$ is passing through either a minimum or a maximum.

1-11. Series Resonant Circuits.- Although the resistance is often only an extremely small fraction of the reactances in a circuit, it may be of great importance because the normal operating value of ω is likely to be such that the reactances may cancel in some of their effects, and the resistance may very largely control the character of the response.

Consider for example the basically important series resonant circuit illustrated in Fig. 1-6. The impedance func-

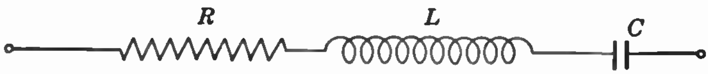


FIG. 1-6. Simple series circuit.

tion $\underline{Z}(\omega)$ has the simple expression

$$\underline{Z}(\omega) = R + j(\omega L - \frac{1}{\omega C}) \text{ complex ohms} \tag{1-26}$$

and the size $Z(\omega)$ of the impedance is

$$Z(\omega) = \sqrt{R^2 + (\omega L - \frac{1}{\omega C})^2} \text{ ohms.} \tag{1-27}$$

Consider that L and C are held constant. Z will be a minimum when $\omega L = 1/\omega C$, or $\omega = 1/\sqrt{LC}$, which will be called ω_1. This is true for any value of R. When ω_1 has this critical value, it is R and R alone which determines and in fact comprises the impedance of the entire circuit. In graphing the function, it will be more convenient to show the reciprocal $Y(\omega)$ of the impedance. This is shown in Fig. 1-7, for a number of values of resistance. At low frequencies, the capacitive reactance is the controlling factor, while at high frequencies the inductive reactance controls. At the critical frequency the admittance is equal to 1/R, and has a maximum value.

The variation of the phase angle of $\underline{Y}(\omega)$ with ω is shown in Fig. 1-8. When the resistance is zero, the admittance function undergoes a sudden jump of 180 degrees, from +90 to -90, as ω is increased and passes through the resonance value $\omega_1 = 1/\sqrt{LC}$.

Applications of the series resonant circuit are obvious and well known. A very low voltage of angular velocity ω_1, when applied to the terminals of the resonant circuit, will produce a greatly amplified voltage as a drop across the inductive reactance and the same across the capacitive reactance. The ratio of amplification will be $\omega_1 L/R$, and hence is inversely proportional to R, other things being equal. One of these reactive drops may be used to energize the grid of an amplifier or other vacuum tube.

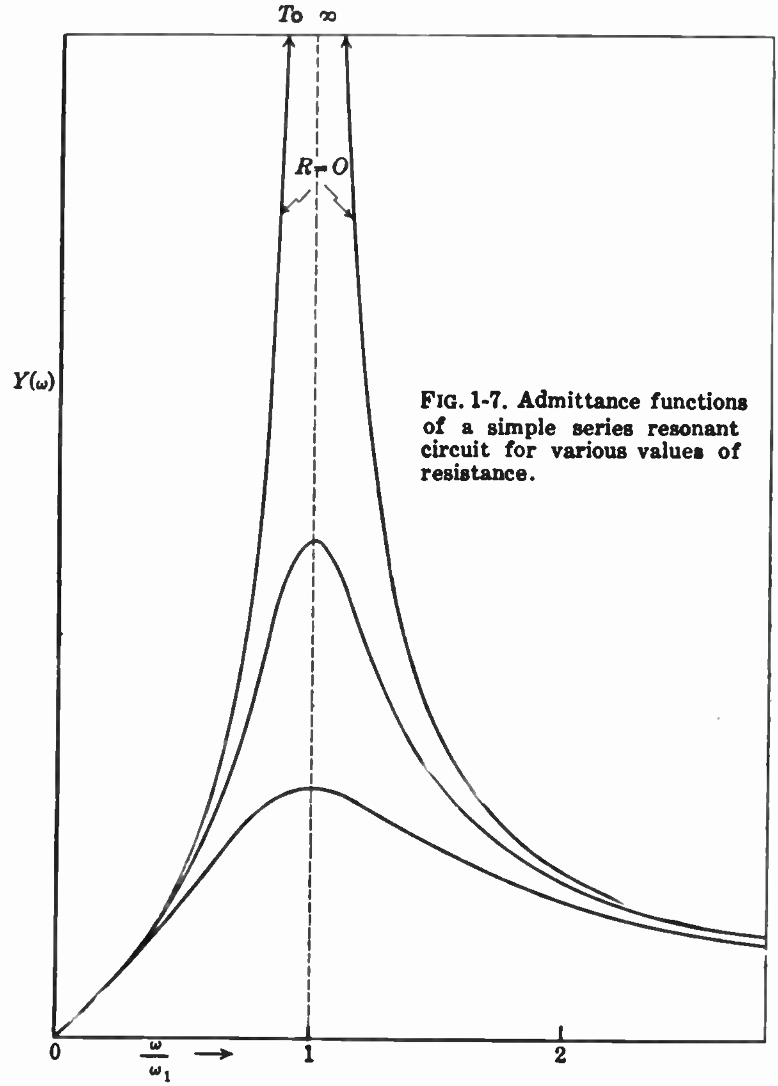


FIG. 1-7. Admittance functions of a simple series resonant circuit for various values of resistance.

The steepness with which the resonance curve $Y(\omega)$ falls from its peak value at $\omega = \omega_1$ is of important practical interest. Let us designate Q as the ratio $\omega L/R$. Very little error is introduced in analyzing the shape of the resonance peak if Q in that vicinity is considered as a constant and equal to $\omega_1 L/R$. With this approximation, it is easy to show that the width of the resonance curve between the "half-power" points is equal to

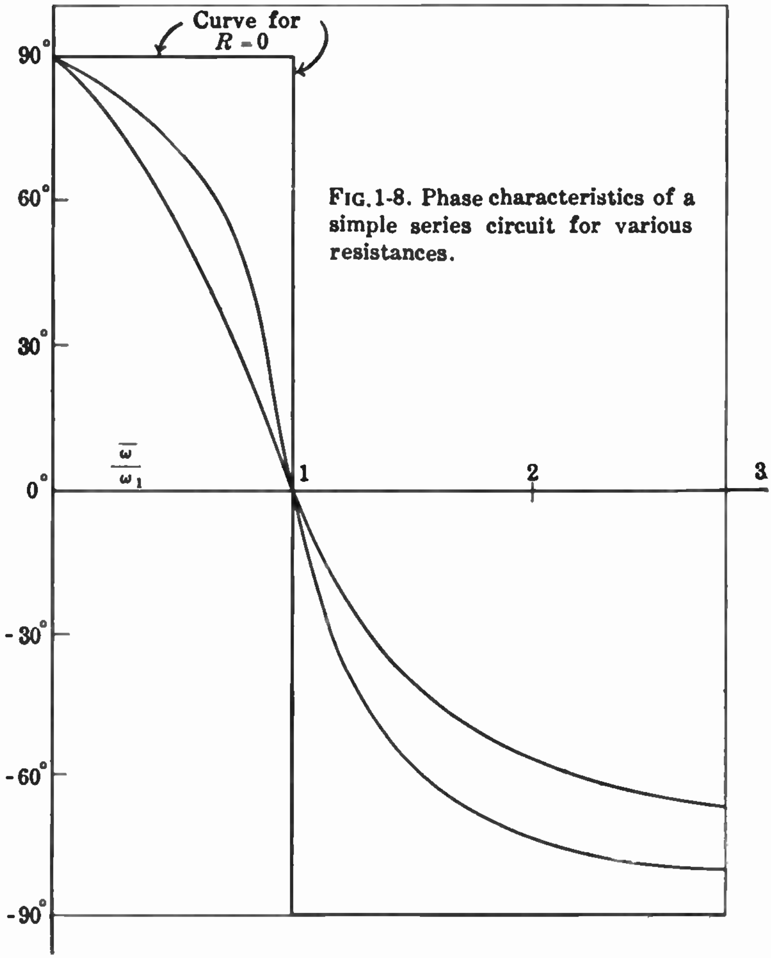


FIG. 1-8. Phase characteristics of a simple series circuit for various resistances.

the fraction 1/Q of the resonance angular frequency ω_1. In other words, when $\omega = \omega_1(1 - 1/2Q)$,

$$\underline{Z} = R - jR \tag{1-28}$$

and when $\omega = \omega_1(1 + 1/2Q)$,

$$\underline{Z} = R + jR. \tag{1-29}$$

In both cases the size of the admittance is $1/\sqrt{2}$ of its peak value. For a fixed impressed voltage the current would of course have $1/\sqrt{2}$ of its resonance peak value, and the power represented would be just one-half its peak value. This approximate rule is very close so long as Q is large, and since it is usually at least 100, no appreciable error is involved.

Two of the curves on each of the Figs. 1-7 and 1-8 have been plotted for very low values of Q, this selection having been made in order to get the entire curve on the sheet. The typical resonance curve as actually used in tuned circuits would have from 25 to 100 or more times the peak values shown in the lower $Y(\omega)$ curves, and the phase angle curves would lie much nearer the curve for $R = 0$. Smith[4] gives some "universal" resonance curves (both Y and phase angle).

Ideal Admittance Characteristics. If a sine-wave voltage $e = E \cos \omega t$ impressed on the input terminals of a linear constant-parameter network produces a current i at the output terminals, i will be a sine wave of amplitude I and phase angle θ determined by the transfer admittance of the network $\underline{Y} = Y\angle\theta$

$$e = E \cos \omega t$$

$$i = EY \cos (\omega t + \theta) = I \cos (\omega t + \theta).$$

If now the input voltage e contains numerous sine-wave components of different frequencies, there will be a component of i corresponding to each component of e. In order that the wave shape of i (total) be the same as that of e, the magnitude of each component of i must be the same multiple of the magnitude of the corresponding component of e as is any other (hence Y must be constant over the range of frequencies concerned) and the phase shifts must keep the relative phases of the components unaltered. This requires that $\theta = -\omega t_d$ where t_d is a constant, and consequently each component of i will be of the form

$$EY \cos \omega(t - t_d)$$

where E and Y are the magnitude of the component and the magnitude of the transfer admittance of angular frequency ω. Thus an ideal characteristic from the point of view of maintaining wave shape is $\underline{Y} = Y_o\angle{-\omega t_d}$ where Y_o and t_d are constants. None of the characteristics shown so far are of this nature. Figure 1-23 shows low-pass filter characteristics which are ideal from the steady-state viewpoint. The quantity t_d is called the delay time, since $EY \cos \omega(t - t_d)$ is "delayed" (lags) with respect to $E \cos \omega t$.

Where wave shape is important, the considerations outlined above are also important, although it is shown later that these steady-state conditions do not serve completely in non-steady-state cases.

1-12. Parallel Resonant Circuits.- A parallel resonant circuit is shown in Fig. 1-9. An inductance element L must

4. Smith, F. Langford, Radiotron Designer's Handbook, third ed., p. 129.

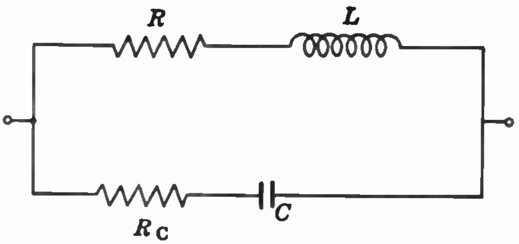


FIG. 1-9. Parallel resonant circuit.

inevitably have some resistance associated with it, and this is shown as R_L. The parallel capacitive element C of course will have some resistance R_C also, but in general it may easily be reduced to an insignificant fraction of that of the other branch. If this is done there is no appreciable error involved in ignoring the condenser resistance altogether. The impedance $\underline{Z}$ of the two individual impedances $\underline{Z}_L$ and $\underline{Z}_C$ in parallel is

$$\underline{Z} = \frac{\underline{Z}_L \underline{Z}_C}{\underline{Z}_L + \underline{Z}_C} \tag{1-30}$$

When R_L is small, there are four distinct frequencies at or near resonance which are very close together and which have special significance. Sometimes they are confused with one another. The following definitions will distinguish among them.

The natural angular frequency ω_n of oscillation of the loop circuit L, R_L, C of Fig. 1-9 is given by the expression

$$\omega_n = \sqrt{\frac{1}{LC} - \frac{R_L^2}{4L^2}}. \tag{1-31}$$

This indicates the frequency of the alternating component of the transient current which would flow in the circuit shown, and is associated also with an exponential decrement factor $e^{-Rt/2L}$.

An angular frequency ω_1 may be defined so that the inductive reactance $\omega_1 L$ and the capacitive reactance $1/\omega_1 C$ are made equal, whence

$$\omega_1 = \sqrt{\frac{1}{LC}}. \tag{1-32}$$

A third definition may represent an angular velocity ω_m at which the impedance of the parallel circuit, as expressed in (1-30), is a maximum. This expression is too complicated to be of much practical utility, and is very nearly equal to ω_1 so long as $\omega_1 L/R$ (i.e., Q) is large.

A fourth definition, ω_o, represents an angular frequency at which the phase angle between total current through, and voltage across, the parallel connection is zero. This is given by

$$\omega_o = \sqrt{\frac{1}{LC}}\sqrt{\frac{L - R_L^2C}{L - R_C^2C}}. \tag{1-33}$$

Because of its simplicity, the expression for ω_1 as given in (1-32) will be used in the further analysis of the circuit.

At angular velocity ω_1, when $\omega L = 1/\omega C$, the denominator of (1-30) reduces to $R_L + R_C$. Under this condition, also, $Z_L = Z_C$ approximately, and so the numerator is $(\omega_1 L)^2$ very nearly. It does not change rapidly with ω, but the denominator does change rapidly with ω if Q is large. The entire function is seen to be, then, the product of a quantity $\underline{Z}_L\underline{Z}_C$ which is approximately constant near the angular velocity ω_1, multiplied by the admittance $1/(\underline{Z}_L + \underline{Z}_C)$ of the <u>series</u> circuit around the loop, which is the exact function whose size is plotted in Fig. 1-7. The impedance function $Z(\omega)$ for the parallel resonant circuit is seen therefore to have a shape near resonance which is the same as that of the admittance function $Y(\omega)$ for the series resonant circuit.

At resonance, the parallel circuit has an impedance which is approximately a pure resistance having a magnitude equal to

$$\frac{(L\omega_1)^2}{R_L + R_C} = QX_L = Q^2R_L \tag{1-34}$$

where Q is $\omega_1 L/R_L$ and X_L is $\omega_1 L$, and $R_L \gg R_C$.

The width of the resonance curve at the half-power points is again equal approximately to the fraction 1/Q of the resonance angular velocity (or frequency), so long as Q is reasonably high, just as in the series resonant circuit.

It is seen from eq. (1-34) that a parallel resonant circuit with high Q develops a very high equivalent resistance across its terminals, much higher (by a factor Q) than the total impedance of either branch. This feature is of value in matching the high impedance of some sources, to obtain maximum power transfer.

The action of the series resonant circuit was seen to be similar to that of a voltage amplifier with an amplification factor of Q at resonance. The parallel resonant circuit, conversely, acts as a current amplifier and produces a circulating current around its loop Q times as large as the line current.

In dealing with resonators to be used in place of tuned circuits at very high frequencies (Ch. 10), it is not convenient to determine directly the R, L, and C of such devices. Instead, there are used the resonance angular frequency $\omega_r = 1/\sqrt{LC}$, the Q at resonance

$$Q_r = \frac{\omega_r L}{R} = \frac{1}{R}\sqrt{\frac{L}{C}}$$

and the input impedance at resonance which, assuming all resistances to be small, is

$$R_e = \frac{L}{RC}.$$

It is seen that these three quantities can specify the circuit of Fig. 1-9 just as well as the usual R, L, C.

1-13. The Transformer.- The basic transformer circuit is shown in Fig. 1-10. It consists of two coils each having self-inductance, and with a mutual inductance between them. The

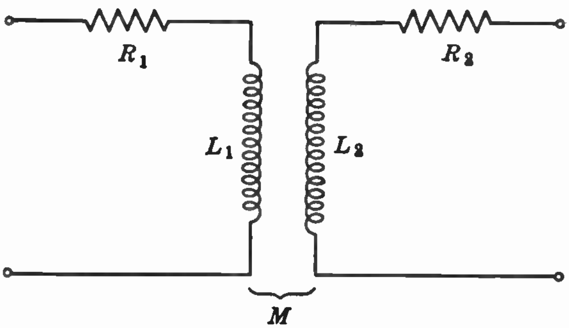


FIG. 1-10. Basic (low-frequency) transformer circuit.

circuit equations, based on positive directions of $\underline{I}_1$ and $\underline{I}_2$ both producing magnetic fields in the same direction in the core, are

$$\underline{E}_1 = (R_1 + j\omega L_1)\underline{I}_1 + j\omega M\underline{I}_2; \tag{1-35}$$

$$\underline{E}_2 = (R_2 + j\omega L_2)\underline{I}_2 + j\omega M\underline{I}_1. \tag{1-36}$$

Somewhat more generally we may write

$$\underline{E}_1 = \underline{Z}_1\underline{I}_1 + j\omega M\underline{I}_2 \tag{1-37}$$

$$0 = \underline{Z}_2\underline{I}_2 + j\omega M\underline{I}_1 \tag{1-38}$$

in which the output voltage $\underline{E}_2$ is supposed connected to some impedance which in series with the transformer secondary impedance produces a total impedance $\underline{Z}_2$. We may eliminate $\underline{I}_2$ between the two equations (1-37) and (1-38) and arrive at an input impedance $\underline{Z}_{11}$ equal to

$$\underline{Z}_{11} = \frac{\underline{E}_1}{\underline{I}_1} = \underline{Z}_1 + \frac{\omega^2 M^2}{\underline{Z}_2} \quad \text{complex ohms.} \tag{1-39}$$

Eliminating $\underline{I}_1$ between them enables us to solve for $\underline{I}_2$, from which the transfer impedance $\underline{Z}_{12}$ is readily obtained as

$$\underline{Z}_{12} = \frac{\underline{E}_1}{\underline{I}_2} = j\left(\omega M + \frac{\underline{Z}_1\underline{Z}_2}{\omega M}\right) \quad \text{complex ohms.} \tag{1-40}$$

By rationalizing and separating reals from imaginaries, the resistance and reactance components of the input and transfer impedances are obtained and in these terms we find:

$$\underline{Z}_{11} = \left[R_1 + \frac{\omega^2 M^2 R_2}{Z_2^2}\right] + j\left[X_1 - \frac{\omega^2 M^2 X_2}{Z_2^2}\right] \quad \text{complex ohms} \quad (1\text{-}41)$$

$$\underline{Z}_{12} = \left[-\frac{R_1 X_2 + X_1 R_2}{\omega M}\right] + j\left[\frac{\omega^2 M^2 + R_1 R_2 - X_1 X_2}{\omega M}\right] \text{complex ohms} \quad (1\text{-}42)$$

It is seen from (1-41) that the input impedance on the primary side is increased by a complex quantity $\frac{\omega^2 M^2 R_2}{Z_2^2} - j\frac{\omega^2 M^2 X_2}{Z_2^2}$, sometimes called the "transferred" or "transformed" impedance. (Transferred impedance should not be confused with "transfer impedance" mentioned in footnote 3 of this chapter.)

The highest possible value for the mutual inductance M in a coupled circuit is $\sqrt{L_1 L_2}$, and this upper limit could be reached only theoretically if all the magnetic flux linking the primary winding also linked every turn of the secondary winding. The ratio $M/\sqrt{L_1 L_2}$ is called the coupling coefficient, with symbol k. In order for k to be nearly equal to unity, it is necessary to provide a transformer core of magnetic material, so as to reduce the amount of leakage flux, or flux which links one winding and not the other. The magnetic material may be thin steel laminations or wires, permalloy, or powdered iron. Core losses take place in these materials, however, and these are larger at high frequencies. The coefficient k seldom exceeds 0.95 even in the best transformers.

1-14. The Resistance-Capacitance Network.- Simple resistance-capacitance networks are in common use in communication circuits; for example, in the coupling between stages of an audio-frequency or video-frequency amplifier. While the actual connections may be very simple; the equivalent circuits which have to be considered for extremely low and high frequencies may be much more elaborate. For example, Fig. 1-11 illustrates a simple resistance-capacitance coupling network[5] comprising only three actual circuit elements, namely a coupling resistance R_c, a coupling condenser C_c, and an output-tube grid-leak resistance R_{go}. The equivalent circuit, also shown in the same figure, includes six elements, the added ones being the plate resistance r_{pi} of the input or left-hand tube, the plate (and lead) capacitance C_{pi} of the input tube to ground, and the grid (and lead) capacitance C_{go} of the output tube. The direct-current supplies are not shown. Over an intermediate range of frequencies, the coupling capacitance C_c has a negligible reactance,

5. See Sec. 3-6 and Fig. 3-10; also Sec. 3-19, part A.

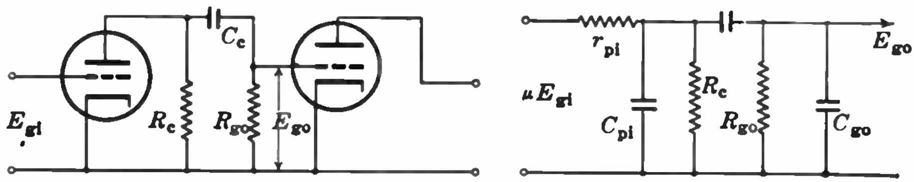


FIG. 1-11. Physical and equivalent circuit of a simple resistance-capacitance coupling network. The subscript *c* on R_c and C_c indicates that these are part of the coupling network; subscript i indicates input tube and subscript o refers to output tube.

and the grid-leak resistance R_{go} will draw a negligibly small current. The effect of the plate and grid capacitances is also negligible over this intermediate range. The value of E_{go} over this range is therefore given very closely by the expression

$$E_{go} = \mu E_{gi} \frac{R_c}{r_{pi} + R_c} \quad \text{volts.} \tag{1-43}$$

At very low frequencies, the plate and grid capacitances will be of negligible effect, but the current drawn by the grid resistance R_{go} will be of importance in causing a voltage drop through the coupling capacitance C_c, whose reactance is inversely proportional to frequency. We may use the actual circuit as shown in Fig. 1-11, with the inclusion of the plate resistance r_{pi}, but this has the disadvantage of involving a series-parallel combination. It is advantageous to use the equivalent constant-current source described in Sec. 1-8, and to obtain thereby the low-frequency equivalent circuit of Fig. 1-12. Since r_{pi} and R_c are in parallel, they have an equivalent resistance which may be designated by a single symbol R_a, where

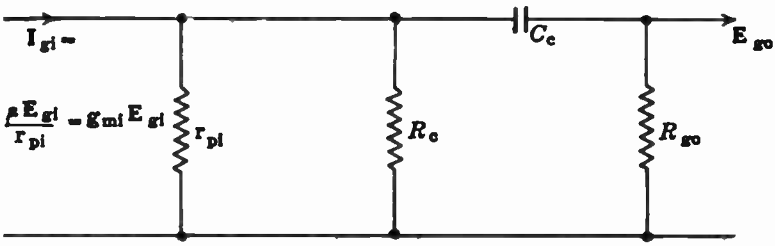


FIG. 1-12. Low-frequency equivalent circuit of resistance-capacitance coupling network.

$$R_a = \frac{R_c r_{pi}}{R_c + r_{pi}} \qquad (1\text{-}44)$$

The output voltage E_{go} is merely the drop across R_{go}, and a simple analysis shows that

$$\underline{E}_{go} = \underline{E}_{gi} g_{mi} R_a \frac{R_{go}}{R_a + R_{go} + 1/j\omega C_c}. \qquad (1\text{-}45)$$

R_a is usually much smaller than R_{go}, and in the denominator it may therefore be dropped out. The ratio of the sizes of $\underline{E}_{go}$ and $\underline{E}_{gi}$ may be written, to this approximation, as

$$\frac{E_{go}}{E_{gi}} = \frac{g_{mi} R_a R_{go}}{\sqrt{R_{go}^2 + X_c^2}}. \qquad (1\text{-}46)$$

The phase angle between $\underline{E}_{go}$ and $\underline{E}_{gi}$ is $\tan^{-1} X_c/R_g$. There is a reduction in $\underline{E}_{go}/\underline{E}_{gi}$ to 71 per cent of its intermediate-frequency value when $X_{Cgo} = R_{go}$.

At very high frequencies the coupling condenser C_c has only a negligible effect, being equivalent to a short circuit, and so all the remaining elements are in parallel, if the constant-current source equivalent of Fig. 1-12 is used. The stray capacitances, plate and grid, have to be considered, and the equivalent circuit is that shown in Fig. 1-13. Here R_e is the

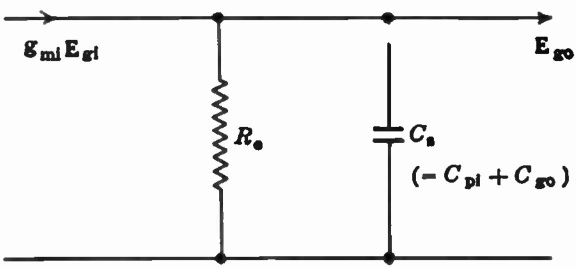


Fig. 1-13. High-frequency equivalent circuit of resistance-capacitance coupling network.

equivalent resistance of r_{pi}, R_c and R_{go} in parallel. The output voltage is merely the impedance drop across the parallel impedances R_e and $1/j\omega C_s$, due to current $E_{gi} g_{mi}$ and is readily shown to have a size given by

$$\frac{E_{go}}{E_{gi}} = \frac{g_{mi} R_e}{\sqrt{1 + \omega^2 C_s^2 R_e^2}} \qquad (1\text{-}47)$$

and a phase angle relative to $\underline{E}_{gi}$ of $-\tan^{-1} \omega C_s R_e$. The size of E_{go}/E_{gi} drops to 71 per cent of its value at intermediate frequencies when $X_s = R_e$.

The network has been analyzed by the use of three different equivalent circuits, each appropriate to a restricted range of frequencies. The advantage of this procedure becomes apparent when the expression is considered which gives the complete response over the entire frequency range. It is

$$\frac{\underline{E}_{go}}{\underline{E}_{gi}} = \frac{\mu \underline{Z}_1 \underline{Z}_2}{\underline{Z}_1\underline{Z}_2 + r_{pi}(\underline{Z}_1 + \underline{Z}_2) + (\underline{Z}_2 + r_{pi})/j\omega C_c}, \tag{1-48}$$

where

$$\underline{Z}_1 = \frac{R_{go}(1 - j\omega C_{go}R_{go})}{1 + \omega^2 C_{go}^2 R_{go}^2}, \tag{1-49}$$

$$\underline{Z}_2 = \frac{R_c(1 - j\omega C_{pi}R_c)}{1 + \omega^2 C_{pi}^2 R_c^2}. \tag{1-50}$$

A much more complicated equation would be required to state explicitly the ratio of sizes of $\underline{E}_{go}$ and $\underline{E}_{gi}$, or the relative phase angle.

1-15. Definition of Q.- The "Q" of a simple inductance coil is defined as the ratio of $\omega L/R$ or X_L/R of the coil. For a condenser, the Q is again defined as X/R, but here of course this is $1/\omega CR$, R in both cases being considered exclusively as a series circuit parameter.

At very high frequencies, and in some other cases also, a more general definition of Q is required. This is

$$Q = \frac{\text{volt-amperes}}{\text{watts dissipated}} \tag{1-51}$$

for the circuit element whose Q is to be evaluated. Q is seen to be a figure of merit. Typical values of Q for coils lie between 100 and 300, and 50 to several thousand for condensers, depending on the insulation and the frequency primarily. The Q of a resonant circuit is equal to the product of the time constant of the envelope of the transient oscillation, multiplied by the natural angular velocity. This definition applies as well to hollow resonators.

The definition of Q in (1-51) may be seen from circuit considerations thus: assume a sine-wave current of rms value I (or peak value $\sqrt{2}$ I), and frequency $f = 1/T$; then Q may be written

$$Q = \frac{\omega L}{R} = \frac{(\omega L I) I}{R I^2} = \frac{E I}{R I^2}$$

which is (1-51). Another way of looking at Q, which is important at very high frequencies, is shown by

$$Q = \frac{\omega L}{R} = \frac{2\pi}{(\frac{1}{2}T)} \cdot \frac{\frac{1}{2}LI^2}{RI^2} = 2\pi \times \frac{\text{energy stored per half cycle}}{\text{energy dissipated per half cycle}}$$

1-16. Tuned Coupled Circuits.- Tuned coupled circuits are used widely in interstage coupling, and in band-pass filter coupling. The coupling may be accomplished by mutual inductance, or by the use of a common self-inductance or common capacitance. Either one or both of the coupled circuits may be tuned, or a chain of more than two coupled circuits may have tuning in each circuit.

In the tuned coupled circuit[6] illustrated in Fig. 1-14, coupling is by the mutual inductance M, and there is a series

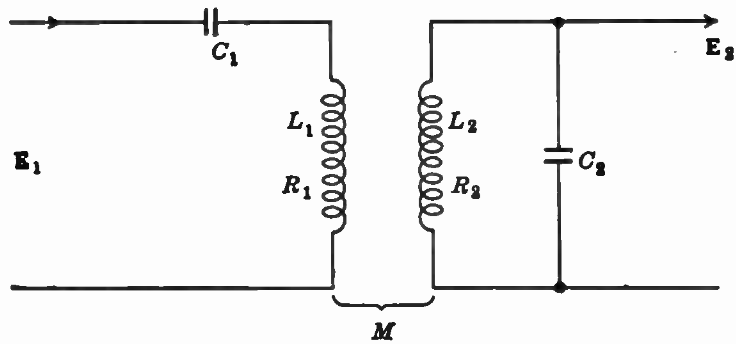


FIG. 1-14. Tuned coupled circuits.

tuning condenser on each side. Such an arrangement is sometimes used in the coupling of band-pass circuits.

There have already been developed, in eqs. (1-40) and (1-42), expressions for the transfer impedance $\underline{Z}_{12}$ of a mutually coupled circuit. An expression for the output voltage $\underline{E}_2$ over the entire frequency range, and for any degree of tuning on the two sides, is obtained from the relation $\underline{E}_2 = \underline{E}_1/j\omega C_2\underline{Z}_{12}$. However, it will be much more convenient to develop an approximate expression for $\underline{E}_2$ which will be good over a restricted range near the resonance frequency.

The assumption is made that the primary and secondary are both tuned to the same frequency: that is,

$$L_1C_1 = L_2C_2 \tag{1-52}$$

Let the resonance angular velocity be defined as

$$\omega_0 = 1/\sqrt{L_1C_1} = 1/\sqrt{L_2C_2}. \tag{1-53}$$

This general coupled circuit may be simplified by considering it as the equivalent of one in which $L_2 = L_1$, and having

6. See Sec. 3-18, paragraph beginning "Doubly-Tuned Transformer Coupling."

an ideal transformer in the secondary circuit to correct the voltage ratio. The ideal transformer would have a ratio $\sqrt{L_1/L_2}$. We may then without loss of generality analyze a circuit whose primary has parameters R_1, L, and C, mutual M, and secondary R_2, L, and C.

In the range near resonance, which is of primary interest, $\omega_o L \gg \omega L - 1/\omega C$, the total reactance in the primary or secondary. For a small change in angular velocity from ω_o to ω there is an approximately equal change in the inductive and capacitive reactance, so that the total reactance on either side is $2L(\omega - \omega_o)$. From eq. (1-42) we may write, for conditions near resonance,

$$\frac{E_2}{E_1} = \frac{j\omega MR_2}{j[R_1R_2 + \omega^2M^2 - 4(\omega - \omega_o)^2L^2] - 2(\omega - \omega_o)L(R_1 + R_2)}. \quad (1\text{-}54)$$

The mutual reactance ωM in the numerator may without much error be considered as equal to $\omega_o M$, a constant, over the restricted range near resonance. In the denominator, however, it has to be considered as a variable. The response E_2/E_1 is inversely proportional to the size of the denominator of (1-54), since the numerator is essentially constant.

As the value of M is varied, the other parameters remaining fixed, the response curve E_2/E_1, plotted as a function of frequency, takes on different shapes. When M is large, nearly equal to L, there are two distinct and well-separated peaks, with a depression in between. As M is reduced, the peaks converge, maintaining approximately the same height. The depth of the depression decreases, until finally when the mutual reactance $\omega_o M$ reaches a value

$$\omega_o M = \frac{\sqrt{R_1^2 + R_2^2}}{\sqrt{2}}, \quad (1\text{-}55)$$

both maxima and minimum are coincident at $\omega = \omega_o$.

If $R_1 = R_2 = R$ then $\omega_o M = R$. For still lower values of mutual reactance, the gain drops materially, there being but a single peak to the curve.

If the primary and secondary circuits have Q's respectively Q_1 and Q_2, then the critical coefficient of coupling k as defined above is

$$k = \frac{1}{\sqrt{2}}\sqrt{\frac{1}{Q_1^2} + \frac{1}{Q_2^2}}, \quad (1\text{-}56)$$

or if $Q_1 = Q_2$,

$$k = \frac{1}{Q}. \quad (1\text{-}57)$$

Since the Q of a coil may easily be two or three hundred, this indicates a very loose coupling.

The critical coupling gives a desirable characteristic because it combines nearly maximum response, a fairly wide flat top to the resonance curve, and absence of frequency distortion due to the presence of two separate peaks (see Fig. 3-47 of Ch. 3). With critical coupling, the voltage response is down to 71 per cent of its maximum when

$$\omega - \omega_o = \pm \frac{R_1 + R_2}{\sqrt{2}L} . \tag{1-58}$$

or

$$\frac{\omega}{\omega_o} - 1 = \pm \frac{1}{\sqrt{2}}\left(\frac{1}{Q_1} + \frac{1}{Q_2}\right). \tag{1-59}$$

Other types of coupled circuit to which the same analysis and formulas apply include those with a common capacitance branch or a common self-inductance branch.

1-17. Broad-Band or Compensating Coupling Networks.- From eq. (1-47), which states the gain or ratio of sizes of $\underline{E}_{go}$ to $\underline{E}_{gi}$ in a resistance-capacitance coupling network, it has been seen that there is a drop to the half-power point, or 0.707 of maximum voltage gain, when the frequency becomes so high that the reactance of the stray shunt capacitance C_s is equal to the equivalent resistance R_e. If we call the angular frequency at which this occurs ω_o,

$$\omega_o = \frac{1}{C_s R_e}, \tag{1-60}$$

and the high-frequency gain may be written

$$\frac{\underline{E}_{go}}{\underline{E}_{gi}} = \frac{g_{mi} R_e}{\sqrt{1 + (\omega/\omega_o)^2}} . \tag{1-61}$$

Since the numerator $g_{mi}R_e$ of (1-61) is essentially constant, the curve of gain will evidently be fairly flat until ω begins to approach ω_o in size, and the flat part may be extended by making ω_o as large as possible. This requires that the capacitance C_s be made as small as practicable, and so special tubes need to be selected or designed with this object in mind. A large transconductance aids in improving the gain all along the curve, but not in extending the range over which the gain is fairly close to constant, except in so far as it affects ω_o.

By inspecting again the high-frequency equivalent circuit of Fig. 1-13, it is readily realized that the cause of the droop in gain around ω_o is the fact that the reactance of C_s is becoming so small that it is draining away most of the hypothetical constant current $E_{gi} g_{mi}$, leaving less to flow through R_e, whose ohmic drop is the output voltage E_{go}.

In order to improve conditions, we should like to increase the reactance of C_s at high frequencies, while not materially lowering it at intermediate frequencies. Optionally, we may increase the impedance of the R_e branch at high frequencies so that its smaller current will still produce a relatively large voltage drop. It is necessary to keep the circuit simple, because for broad-band video amplifiers, to work up to about 4,000,000 cycles per second, the distributed capacitance effects are very important. The size of C_s, which needs to be kept to a minimum, would increase greatly if an elaborate network were used to approximate a constant impedance with varying frequency. It is therefore the practice in some video amplifiers to add a small inductance in series with the coupling resistance R_e producing an equivalent circuit approximating that shown in Fig. 1-15. Based on this equivalent circuit having an

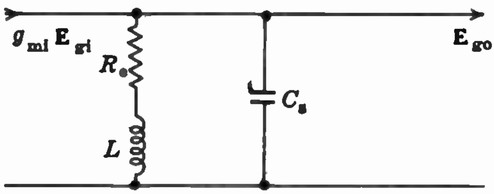


FIG. 1-15. High-frequency equivalent of resistance-capacitance-coupling network, with inductance L for high-frequency response correction.

impedance $\underline{Z}$ of its two branches in parallel, the gain is evidently

$$\frac{E_{go}}{E_{gi}} = g_{mi}Z \qquad (1\text{-}62)$$

where

$$Z = \text{size of } \frac{(R_e + j\omega L)/j\omega C_s}{R_e + j\omega L + 1/j\omega C_s}$$

$$= \frac{\sqrt{\frac{R_e^2}{\omega^2 C_s^2} + \left(\frac{L}{C_s} - \omega^2 L^2 - R_e^2\right)^2}}{\omega C_s\,[R_e^2 + (\omega L - 1/\omega C_s)^2]}. \qquad (1\text{-}63)$$

At low and intermediate frequencies this expression is equal approximately to R_e. The phase angle of $\underline{Z}$ is

$$\tan^{-1}\frac{\frac{L}{C_s} - \omega^2 L^2 - R_e^2}{R_e/\omega C_s}. \qquad (1\text{-}64)$$

By making use of the abbreviations

$$\omega_0 = \frac{1}{R_e C_s}, \qquad (1\text{-}65)$$

$$K = \frac{L}{C_s R_e^2}, \tag{1-66}$$

the ratio of Z to its intermediate-frequency value is expressed as

$$\frac{Z}{R_e} = \frac{E_{go}}{E_{go\ int}} = \sqrt{\frac{1 + K^2\omega^2/\omega_o^2}{1 - (2K - 1)\omega^2/\omega_o^2 + K^2\omega^4/\omega_o^4}} \tag{1-67}$$

and the phase angle is

$$\tan^{-1} [(K - 1)\omega/\omega_o - K^2\omega^3/\omega_o^3]. \tag{1-68}$$

To find the condition which will make the gain at ω_o the same as at lower frequencies, we may equate (1-67) to unity while making ω/ω_o also unity, and solve for K, which must be equal to 1/2. This does not of course indicate a flat response characteristic out to $\omega = \omega_o$. Actually, the gain rises to a small peak for values of ω somewhat less than ω_o, and then declines continuously as ω is increased, passing through the reference value when $\omega = \omega_o$.

The relative phase delay when $\omega = \omega_o$ and $K = 1/2$, as compared to a reference value of delay which holds constant over the intermediate frequencies, is, from (1-68),

$$\tan^{-1} \left(- \frac{3}{4}\right) \text{ for } \omega = \omega_o \tag{1-69}$$

$$\tan^{-1} \left(- \frac{1}{20} - \frac{1}{4000}\right) \text{ for } \omega = \omega_o/10 \tag{1-70}$$

and the time delays for the two angular velocities are $-0.643/\omega_o$ and $-0.0502/0.1\omega_o$ second, respectively. Thus it is seen that the difference in delay is $0.141/\omega_o$ second.

Conversely, a value of K may be determined that will yield the same time delay at ω_o as at intermediate frequencies by equating the product of (1-68) and $1/\omega$ for the two values of angular frequency $\omega = \omega_o$ and $\omega = 0.1\omega_o$. This procedure indicates a value of K equal to 0.32. Since the shape of the phase characteristic is important in video amplifiers, this figure must be given some weight, and if ω_o is the upper limit of the desired operating range, a compromise value of K approximately halfway between 0.5 and 0.32 should be used.

Six variations of the high-frequency compensating coupling network are shown in Fig. 1-16.

1-18. Impedance Matching Networks.- It has been shown in Sec. 1-9 that the conditions for maximum power transfer require that the load impedance be made equal to the conjugate of the impedance of the source. This may not always be the most desirable condition, because it usually involves operation at an efficiency of only 50 per cent. It is the criterion in most communication circuits, however, and even when it is not the

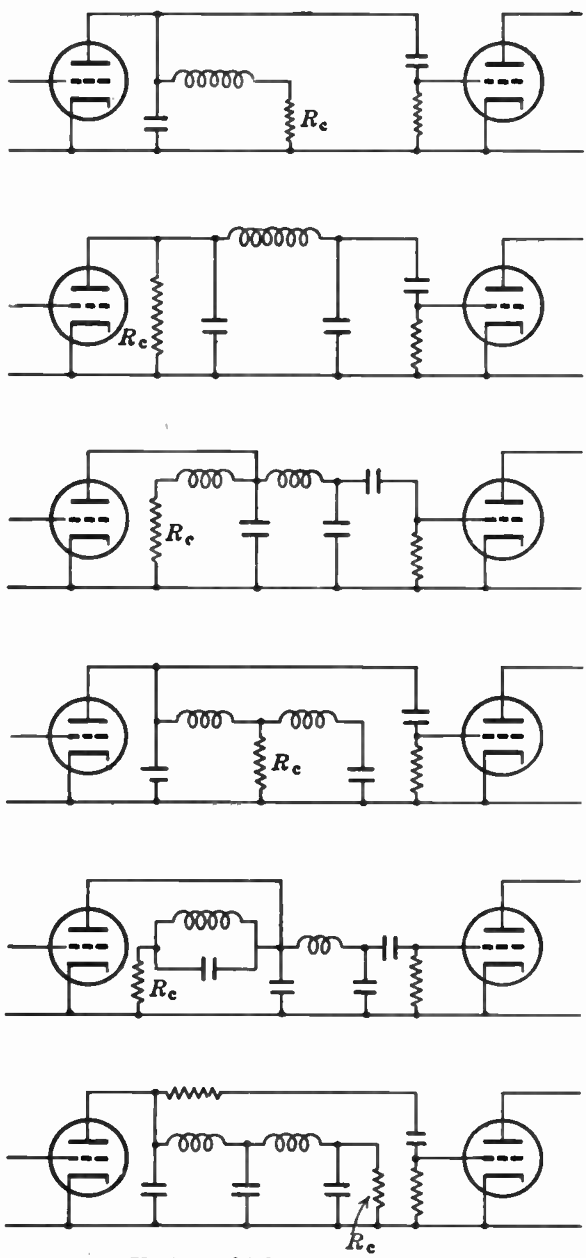


FIG. 1-16. Various high-frequency compensating coupling circuits (D. L. Jaffe, Electronics, April 1942).

sole criterion, it does have some weight in the decision as to what value the load impedance should have.

It is very often necessary for a source to supply power to a load of radically different impedance. Some sort of matching network is then of advantage, which for exact matching should have an input impedance equal to the impedance of the source, and an output impedance (i.e., impedance looking into the matching network from its output terminals) equal to the load impedance.

The simplest and most common circuit for impedance matching is the transformer. If the coefficient of coupling is nearly unity, the ratio of voltage transformation is approximately the same as the ratio of turns, and the impedance ratic varies as the square of the number of turns. Thus, to supply a 100-ohm load from a 900-ohm source, a 3-to-1 transformer should be used. Considering it as an ideal transformer, that is, one without losses or leakage reactance, the impedance looking toward the load from the high-tension primary would be 900 ohms, a perfect match for the source. Also, the impedance looking toward the source from the low-tension secondary would be 100 ohms, a perfect match for the load.

The coupling transformer may be of the autotransformer type, with the two low-tension terminals taken from one end of the complete winding and from some tap point along the winding, with the high-tension terminals coming from the two ends (usually) of the winding.

Impedance matching over a restricted narrow frequency range may be accomplished by taking taps from the circuit elements of a resonant circuit, and by means of connections made to pieces of transmission line having distributed parameters. This method of matching will be discussed in connection with the theory of transmission lines, in Ch. 11. Equivalent "lumpy" networks may also be used for the same purpose.

1-19. Simple Filter Circuits.- Consider the uniform recurring network illustrated in Fig. 1-17.

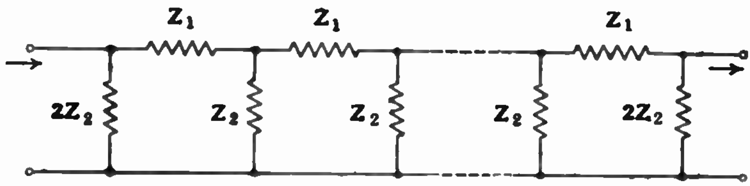


FIG. 1-17. Ladder network.

All the series impedance elements are the same, each equal to $\underline{Z}_1$ complex ohms, and all the shunt elements, with the exception of those at each end, are equal each to $\underline{Z}_2$. This arrangement is the equivalent of a number of π sections connected in series, each π having leg impedances of $2\underline{Z}_2$ on each side. It may be shown that this type of structure has very useful properties as a filter, in attenuating greatly certain frequencies and passing others with relatively little attenuation.

Suppose there are an infinite number of the π sections, of which one is shown in Fig. 1-18, connected in series. Let

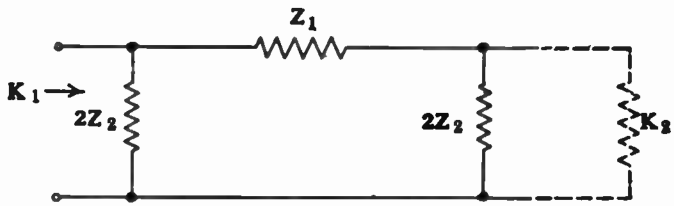


FIG. 1-18. Elementary π- section of mid-shunt filter.

the input impedance of the infinite ladder be $\underline{K}_2$ complex ohms. Then if the single elementary π section is given a load of impedance $\underline{K}_2$ at one end, the impedance looking into the other end must be $\underline{K}_2$ also. We may write then

$$\underline{K}_2 = \frac{\left(\dfrac{2\underline{Z}_2\,\underline{K}_2}{2\underline{Z}_2 + \underline{K}_2} + \underline{Z}_1\right) 2\underline{Z}_2}{\dfrac{2\underline{Z}_2\,\underline{K}_2}{2\underline{Z}_2 + \underline{K}_2} + \underline{Z}_1 + 2\underline{Z}_2} \quad \text{complex ohms} \tag{1-71}$$

whence, solving for $\underline{K}_2$ in terms of $\underline{Z}_1$ and $\underline{Z}_2$,

$$\underline{K}_2 = \frac{\underline{Z}_1\underline{Z}_2}{\sqrt{\underline{Z}_1\underline{Z}_2 + \underline{Z}_1^2/4}} \quad \text{complex ohms.} \tag{1-72}$$

If $\underline{Z}_1$ and $\underline{Z}_2$ are so related that $\underline{K}_2$ is a constant, and not a function of frequency, then the filter is called a constant K filter.

If the current is taken as $\underline{I}$ and the voltage $\underline{K}_2\underline{I}$ at the load end of the π section, the corresponding values of current and voltage at the other end are found to be these same values each multiplied by the same factor

$$e^{\underline{\gamma}} = \frac{2\underline{Z}_2 + \underline{K}_2}{2\underline{Z}_2 - \underline{K}_2} \tag{1-73}$$

This equation will serve as a definition of the symbol $\underline{\gamma}$, which is called the propagation constant per section. Also, we set

$$\gamma = \alpha + j\beta \tag{1-74}$$

where α and β are real quantities. Alpha (α) is the attenuation constant per section, and β the phase constant per section. A current or voltage, in passing through any number n sections, will be attenuated by a factor $e^{-n\alpha}$, and will undergo a phase retardation in its direction of travel amounting to $n\beta$ radians.

We shall concern ourselves here with the attenuation constant α, which it must be remembered is a function of frequency. If we substitute back into (1-73) the value of $\underline{K}_2$ from (1-72) and simplify, we have

$$e^{\gamma} = \frac{\sqrt{1 + 4\underline{Z}_2/\underline{Z}_1} + 1}{\sqrt{1 + 4\underline{Z}_2/\underline{Z}_1} - 1} \quad \text{complex numeric.} \tag{1-75}$$

On the assumption that the ladder impedances $\underline{Z}_1$ and $\underline{Z}_2$ are made up of lossless elements, they will be pure imaginary numbers and their ratios real, and either positive or negative. If the twin radicals in (1-75) are zero or imaginary, then the numerator and denominator sizes are identical, although there will be a difference in phase angle. This condition will be characteristic of the pass band or bands of the structure, and evidently holds when

$$\frac{4\underline{Z}_2}{\underline{Z}_1} \leq -1. \tag{1-76}$$

In other words, $4Z_2 \geq Z_1$ in size, and the two reactances must be of opposite sign (one inductive, the other capacitive) in the frequency range which is passed without attenuation.

When the radical is real, the numerator is larger than the denominator, and the circuit has attenuation, but no phase shift with ideal lossless elements.

Actually, of course, the circuit elements will have some loss, but the losses may be kept small enough usually so that their effects are relatively unimportant.

Low-pass Filter. Suppose that the series impedances $\underline{Z}_1$ consist only of inductances of L henrys each, and the shunt impedances $2\underline{Z}_2$ of capacitances C farads each, as shown in Fig. 1-19. Here

$$\underline{Z}_1 = j\omega L \tag{1-77}$$

$$\underline{Z}_2 = 1/j\omega C \tag{1-78}$$

$$4\underline{Z}_2/\underline{Z}_1 = -4/\omega^2 LC \tag{1-79}$$

The critical angular frequency ω_1, from (1-76), will be

$$\omega_1 = \frac{2}{\sqrt{LC}} \tag{1-80}$$

and for values of ω less than this, there is no attenuation (in the absence of resistance or other losses) as shown in the curve

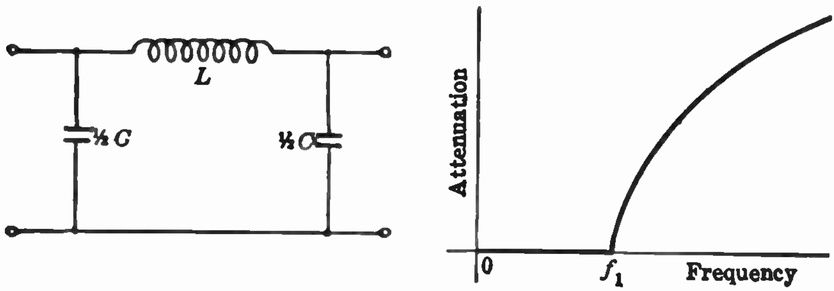


FIG. 1-19. Low-pass filter π-section, and attenuation characteristic.

of attenuation plotted against frequency in Fig. 1-19.

High-pass Filter. If the series $\underline{Z}_1$ is a purely capacitive element ($\underline{Z}_1 = 1/j\omega C$) and each shunt $\underline{Z}_2$ an inductive element ($\underline{Z}_2 = j\omega L$), then

$$4\underline{Z}_2/\underline{Z}_1 = -4\omega^2 LC \qquad (1\text{-}81)$$

and the critical angular frequency ω_1 at which this expression becomes equal to -1 is

$$\omega_1 = \frac{1}{2\sqrt{LC}} \quad \text{radians per second.} \qquad (1\text{-}82)$$

At frequencies greater than the critical frequency, the attenuation would be zero (with lossless elements making up the filter), while at lower frequencies there is attenuation. The attenuation curve of Fig. 1-19 would serve also for the high-pass filter if the frequency scale is used as the reciprocal of the values of f/f_1 indicated for the low-pass characteristic.

Band-pass and band-elimination characteristics are obtained by proper design of the $\underline{Z}_1$ and $\underline{Z}_2$ elements. There is a wide literature on filter design to which the reader must be referred for more detailed analysis of these more complicated realizations, the m-derived, the lattice type, and other structures.

The analysis given here was based on the π section, but the T or the L section would serve equally well. For the T-type, or mid-series terminated filter, the ladder network of Fig. 1-17 would have a series element of $\underline{Z}_1/2$ at each end, all other series elements being $\underline{Z}_1$ and all shunt elements $\underline{Z}_2$. The characteristic impedance $\underline{K}_1$ of this infinite structure would be

$$\underline{K}_1 = \sqrt{\underline{Z}_1\underline{Z}_2 + \underline{Z}_1^2/4}, \qquad (1\text{-}83)$$

and

$$e^{\underline{\gamma}} = \frac{2\underline{K}_1 + \underline{Z}_1}{2\underline{K}_1 - \underline{Z}_1}. \qquad (1\text{-}84)$$

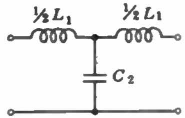
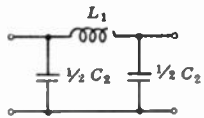
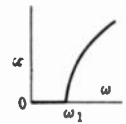
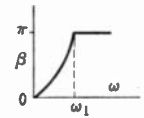
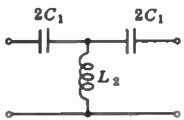
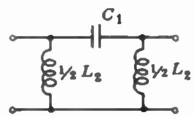
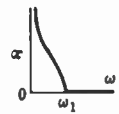
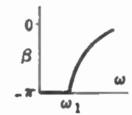
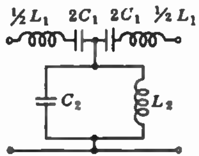
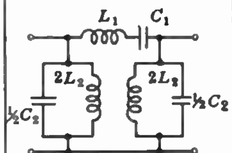
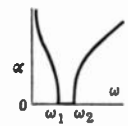
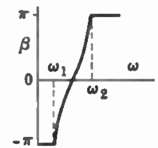
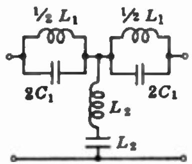
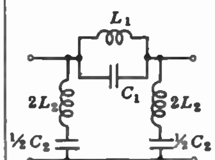
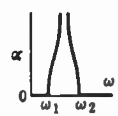
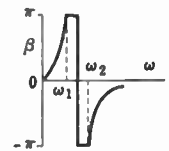
Type	T - section	π - section	Circuit parameters (assuming cutoff frequencies and K are specified)	Attenuation characteristic α	Phase characteristic β
Low-pass			$L_1=2K/\omega_1$ $C_2=2/\omega_1 K$ $K^2=L_1/C_2$; $\omega_1^2=4/L_1C_2$		
High-pass			$C_1=1/2\omega_1 K$ $L_2=K/2\omega_1$ $K^2=L^2/C_1$; $\omega_1^2=1/4L_2C_1$		
Band-pass			$L_1C_1=L_2C_2=1/\omega_1\omega_2$ $L_1=2K/(\omega_2-\omega_1)$ $C_2=2/K(\omega_2-\omega_1)$ $K^2=L_2/C_1=L_1/C_2$		
Band-elimination			$L_1C_1=L_2C_2=1/\omega_1\omega_2$ $C_1=1/2K(\omega_2-\omega_1)$ $L_2=K/2(\omega_2-\omega_1)$ $K^2=L_2/C_1=L_1/C_2$		

FIG. 1-20. Simple constant-K filter sections.

1-20. Change in Impedance Level in Network Design.- When a filter has to work out of a source of one value of impedance into a load of another level, it is possible to accomplish the impedance matching by using unsymmetrical filter terminations. Thus an L-section filter, in which each section has a series impedance $\underline{Z}_1$ and a single-shunt impedance $\underline{Z}_2$ at one end, exhibits a different characteristic impedance depending upon which end is used as the input. Looking into an infinite line of these L sections from a series $\underline{Z}_1$ termination, that is, from the left in Fig. 1-21, we may write for the characteristic impedance $\underline{Z}_3$

$$\underline{Z}_3 = \underline{Z}_1 + \frac{\underline{Z}_3\underline{Z}_2}{\underline{Z}_3 + \underline{Z}_2} \quad \text{complex ohms,} \tag{1-85}$$

whence

$$\underline{Z}_3 = \frac{\underline{Z}_1}{2} \pm \sqrt{\underline{Z}_1\underline{Z}_2 + \underline{Z}_1^2/4} \quad \text{complex ohms} \tag{1-86}$$

The value indicated by the minus sign is not usually physically realizable in a passive network, so only the plus sign will be considered.

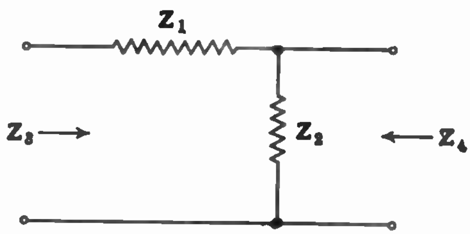


FIG. 1-21. Unsymmetrical filter section.

Now looking into the infinite chain of sections from a shunt termination, such as from right to left in Fig. 1-21, we write for the characteristic impedance $\underline{Z}_4$

$$\underline{Z}_4 = \frac{(\underline{Z}_4 + \underline{Z}_1)\underline{Z}_2}{\underline{Z}_4 + \underline{Z}_1 + \underline{Z}_2} \quad \text{complex ohms,} \tag{1-87}$$

whence

$$\underline{Z}_4 = -\frac{\underline{Z}_1}{2} \pm \sqrt{\underline{Z}_1\underline{Z}_2 + \underline{Z}_1^2/4} \quad \text{complex ohms.} \tag{1-88}$$

Here again the minus sign will be disregarded for the same reason given above. The ratio of the change in impedance level is

$$\frac{\underline{Z}_3}{\underline{Z}_4} = \frac{\dfrac{\underline{Z}_1}{2} + \sqrt{\underline{Z}_1\underline{Z}_2 + \underline{Z}_1^2/4}}{-\dfrac{\underline{Z}_1}{2} + \sqrt{\underline{Z}_1\underline{Z}_2 + \underline{Z}_1^2/4}} \tag{1-89}$$

$$= \frac{\sqrt{1 + 4Z_2/Z_1} + 1}{\sqrt{1 + 4Z_2/Z_1} - 1} \qquad (1\text{-}90)$$

If

$$\frac{4Z_2}{Z_1} > -1 \qquad (1\text{-}91)$$

there is a change in impedance level to the ratio indicated by eq. 1-90. A gradual change or taper in the section constants is sometimes used to correct a small mismatch.

1-21. Non-sinusoidal Waves.- It has been pointed out that, in linear networks, periodic or cyclical voltages[7] having some other shape than a pure sinusoid will have the same total effect in setting up currents or other responses as that obtained by combining the individual responses due to the individual component pure sinusoids into which the non-sinusoidal voltage may be resolved. If the force impressed on the network is considered to be a current rather than a voltage, the same approach may of course be used.

The solution of linear circuit problems with non-sinusoidal waves therefore consists of two steps:

1. Resolve the wave into sinusoidal components.
2. Calculate individually and combine (if desired) the responses to the sinusoidal components.

More often than not, the combination of the responses does not need to be made, because a knowledge of the response at each frequency usually suffices.

1.22. Fourier Series.- Let f(t) be any periodic voltage, or current, or other quantity varying cyclically with the time t with period T (see footnote 7). Then it is possible to expand f(t) as follows:

$$f(t) = \tfrac{1}{2}a_0 + a_1 \cos \omega t + a_2 \cos 2\omega t + \ldots + a_n \cos n\omega t + \ldots$$
$$+ b_1 \sin \omega t + b_2 \sin 2\omega t + \ldots + b_n \sin n\omega t + \ldots \qquad (1\text{-}92)$$

where $\omega \equiv 2\pi/T$. The following are various other forms of the Fourier series, all equivalent to (1-92) and to each other:

7. By a periodic voltage is meant one which mathematically satisfies $f(t) = f(t + nT)$ where n is any integer and T is the (constant) period. In words, this says that the voltage at any time t is the same as at any time an integer number of periods before or after t, and this is true for every point on the cycle. Thus the voltage, or more generally f(t), must repeat itself identically every period.

$$f(t) = \tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos n\omega t + b_n \sin n\omega t)$$

$$= \tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} c_n \cos (n\omega t - \theta_n) \quad \begin{cases} c_n = \sqrt{a_n^2 + b_n^2} \\ \theta_n = \tan^{-1} (b_n/a_n) \end{cases}$$

$$= \sum_{n=-\infty}^{\infty} \underline{d}_n e^{jn\omega t} \quad \begin{cases} \underline{d}_n = \tfrac{1}{2}(a_n - jb_n) \text{ for } n \\ \text{positive, conjugate for} \\ n \text{ negative} \end{cases} \qquad (1\text{-}93)$$

The coefficients a_n and b_n may be determined in the following way. Multiply both sides of (1-92) by cos $n\omega t$, giving

$$f(t) = \tfrac{1}{2}a_0 \cos n\omega t + a_1 \cos \omega t \cos n\omega t + a_2 \cos 2\omega t \cos n\omega t$$
$$+ \ldots + a_n \cos^2 n\omega t + \ldots + b_1 \sin \omega t \cos n\omega t$$
$$+ b_2 \sin 2\omega t \cos n\omega t + \ldots + b_n \sin n\omega t \cos n\omega t + \ldots \qquad (1\text{-}94)$$

Integrating both sides with respect to t over the range from $-\frac{T}{2}$ to $\frac{T}{2}$ reduces every term on the right-hand side to zero except $a_n \cos^2 n\omega t$

$$a_n \int_{-\frac{T}{2}}^{\frac{T}{2}} \cos^2 n\omega t dt = a_n \frac{T}{2} \qquad (1\text{-}95)$$

whence $$a_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) \cos n\omega t dt \quad n=0,1,2,3,.. \qquad (1\text{-}96)$$

and likewise $$b_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) \sin n\omega t dt \quad n=1,2,3,\ldots \qquad (1\text{-}97)$$

Note that $$\underline{d}_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)\, e^{-jn\omega t} dt \quad n=\ldots-2,-1,0,1,2,\ldots \qquad (1\text{-}98)$$

If the function to be represented is the output of a half-wave rectifier, consisting of only the positive halves of a sine wave in which the voltage (or current) is zero from $\omega t = -\pi$ to $\omega t = 0$ and $E_m \sin \omega t$ for ωt from 0 to π, the Fourier series may be shown to be:

$$e = E_m\left[\frac{1}{\pi} + \frac{1}{2} \sin \omega t - \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{\cos 2n\omega t}{4n^2-1}\right]$$
$$= \frac{E_m}{\pi} [1 + 1.57 \sin \omega t - 0.67 \cos 2\omega t - 0.13 \cos 4\omega t$$
$$- 0.06 \cos 6\omega t - \ldots]. \qquad (1\text{-}99)$$

There are no odd harmonics present, with the exception of the fundamental.[8]

The output of a full-wave rectifier consists of all the positive loops and, in addition, the negative loops reversed.

---

8. The fact that the magnitudes of the harmonics fall rapidly with increasing order should be noted. In the design of filters for use with rectifiers this is of some importance (see Ch. 2).

It might be written $e = |E_m \sin \omega t|$, the "absolute value" of $E_m \sin \omega t$. The Fourier series of this is

$$e = E_m\left[\frac{2}{\pi} - \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{\cos 2n\omega t}{4n^2-1}\right]$$

$$= \frac{2}{\pi}E_m[1 - 0.67 \cos 2\omega t - 0.13 \cos 4\omega t - 0.06 \cos 6\omega t \,..] \quad (1\text{-}100)$$

Again there are no odd harmonics present, and in this case no fundamental. The various even harmonics are present in the same percentage of direct potential for both half- and full-wave circuits, but the fundamental ripple, absent in the full-wave circuit, is about 2.5 times as large in the half-wave rectifier as is the second harmonic.

The Fourier series of a repeated pulse of unit height, rectangular shape, and duration t_1, as seen in Fig. 1-22, is

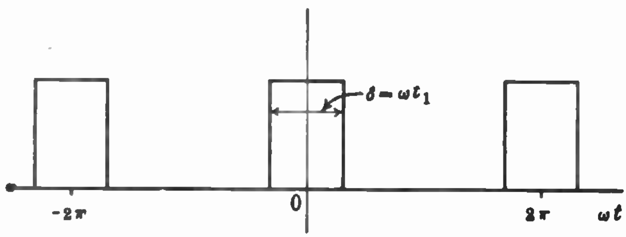


FIG. 1-22. Rectangular pulses, duration t_1, repeated at intervals T ($\omega = 2\pi/T$.)

easily shown, by (1-96) and (1-97), to be

$$f(t) = \frac{\delta}{2\pi} + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\sin n\delta/2}{n}\cos n\omega t, \qquad (101)$$

where $\delta = \omega t_1$. The form of the Fourier series for the same function will be changed by a shift in position with respect to the origin.

1-23. The Fourier Integral.- It is often desired to ascertain the effect on a linear system of an isolated pulse or some other non-recurrent driving force. Obviously such a non-recurrent force function is not expressible as a Fourier series, which from its nature is a periodic function, but it may be expressed analytically by a Fourier integral. Sometimes it is sufficient to assume that the non-periodic driving force is repeated after an interval, and thus to assume that it is periodic, hence resolvable by a Fourier series. But this process is difficult to apply in many of the cases in which it can be used, and often it is not applicable to the problem at hand.

The Fourier integral may be considered the end result of the Fourier series when the period of the "periodic" function to be analyzed is made infinite. Thus a non-periodic function becomes considered the equivalent of a periodic function with infinite period, and for such a function the Fourier series, which is intended for finite periods, changes into the Fourier integral.

Non-periodic driving forces are of considerable importance in present-day ultra-high-frequency work, and the use of the Fourier integral or some equivalent has become a substantial tool for circuit analysis and in handling other types of problems in the field. Many u-h-f systems - for example, television systems - must deal with abrupt, non-recurrent, stimuli.

Recalling eq. (1-93) that the Fourier series may be written

$$f(t) = \sum_{n=-\infty}^{\infty} \underline{d}_n e^{jn\omega t} \qquad (1\text{-}102)$$

and that the $\underline{d}_n$ coefficients are given by eq. (1-98)

$$\underline{d}_n = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)e^{-jn\omega t}dt \qquad (1\text{-}98)$$

it is possible to write one expression for f(t) by substituting for $\underline{d}_n$ in (1-102). Now since the $\underline{d}_n$ coefficients do not depend on t (t is eliminated when the limits are substituted), it is permissible to replace t under the integral sign by any other symbol--say x. Then

$$\underline{d}_n = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(x)e^{-jn\omega x}dx \qquad (1\text{-}103)$$

where f(x) is the same as f(t) except that t has been replaced by x. The use of x in the expression for $\underline{d}_n$ avoids confusion when substituting for $\underline{d}_n$ in (1-102), where the t in the exponential remains in the result; such a substitution gives

$$f(t) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\left[\int_{-\frac{T}{2}}^{\frac{T}{2}} f(x)e^{-j\frac{2\pi n}{T}x}dx\right] e^{j\frac{2\pi n}{T}t} \cdot \frac{2\pi}{T}$$

If now the period T approaches ∞, the quantity $\frac{2\pi}{T}$, heretofore written ω, becomes infinitesimal and may be written dω. Furthermore, n times $\frac{2\pi}{T}$ may, since n becomes as large as we please, be written ω and the sum for n now becomes an integral since unit change in n produces an infinitesimal change in 2πn/T, which is the only factor in which n appears. Hence for an infinite period

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x)e^{-j\omega x}dx\right] e^{j\omega t}d\omega \qquad (1\text{-}104)$$

which may just as well be written

$$f(t) = \int_{-\infty}^{\infty} F(\omega)e^{j\omega t}d\omega \qquad (1\text{-}105)$$

where

$$F(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{-j\omega x}dx. \qquad (1\text{-}106)$$

The derivation of the Fourier integral (1-104) may require on the part of the reader some concentrated attention and a possible review of the definition of a definite integral, and the result as expressed in (1-104) may seem of dubious value since f(t) is expressed in a rather complicated way in terms of itself [f(x)]. But eq. (1-105) shows the heart of the result. If f(t) had been periodic with finite period, then (1-93) states that f(t) might have been expanded into an infinite number of terms of different frequencies, each frequency separated by a finite amount from its nearest neighbor:

$$f(t) = \sum_{n=-\infty}^{\infty} \underline{d}_n e^{jn\omega t} \qquad (1\text{-}93)$$

But f(t) is here taken to be non-periodic; that is, it has an infinite period, and (1-105) states that

$$f(t) = \int_{-\infty}^{\infty} F(\omega)e^{j\omega t}d\omega \qquad (1\text{-}105)$$

Comparing these two expressions, it is seen that the non-periodic f(t) has been expressed as the sum (integral) of an infinite number of components, each of <u>infinitesimal amplitude</u> $F(\omega)d\omega$, <u>infinitely close together</u> (in the integral ω varies <u>continuously</u> whereas in (1-93) $n\omega$ changes by jumps since n is always an integer).

The amplitude of the Fourier series coefficient $\underline{d}_n$ is given by (1-98):

$$\underline{d}_n = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(x)e^{-jn\omega x}dx \qquad (1\text{-}98)$$

whereas the amplitude of each component of f(t) in (1-105) is

$$F(\omega)d\omega = d\left(\frac{1}{T}\right)\int_{-\infty}^{\infty} f(x)e^{-j\omega x}dx \qquad (1\text{-}107)$$

The correspondence is so close that it will not be expanded upon.

The most important use of the Fourier integral for present purposes may be outlined thus:

1. $F(\omega)$ may be obtained from (1-106), and the amplitude $F(\omega)d\omega$ of each component of the driving force f(t) in (1-105) is consequently known. The driving force f(t) is thus resolved and the amplitude of each component determined.

2. If each of the driving force components is multiplied by the transfer admittance $\underline{Y}(\omega)$ which connects the driving force with the effect it produces, the effect due to each component of the driving force can be found.

3. If all the effects determined in (2) are added together, the total response can be determined.

In cryptic form, the rule is: (1) analyze the driving force, (2) multiply each component by $\underline{Y}(\omega)$, (3) synthesize to find the total response.

In mathematical form: given a non-periodic driving force f(t) and a system whose transfer admittance is $\underline{Y}(\omega)$:

1. Determine $F(\omega)$ by (1-105).
2. Multiply $F(\omega)d\omega$ by $\underline{Y}(\omega)$.
3. Determine the response by adding all the component responses thus:

$$\int_{-\infty}^{\infty} F(\omega)\underline{Y}(\omega)e^{j\omega t}\,d\omega \qquad (1\text{-}108)$$

As a simple application, consider the current output (response) due to a suddenly applied emf (driving force) applied to the input terminals of a low-pass filter which on a

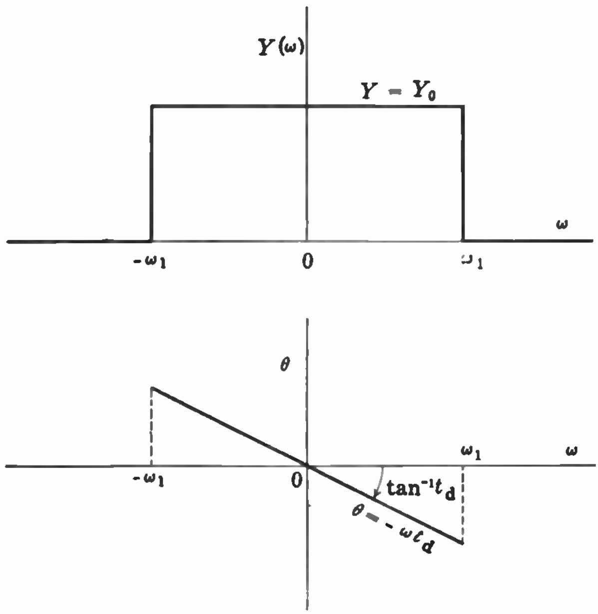


Fig.1-23. Admittance characteristics (magnitude Y and phase angle θ) of a low-pass filter which is ideal from a steady-state point of view.

steady-state basis has an ideal transfer admittance characteristic[9] (Fig. 1-23). The input emf is

$$e = 0 \text{ for } t < 0; \quad e = E \text{ (constant) for } t > 0 \tag{1-109}$$

and is pictured in Fig. 1-24. Such a voltage might be applied by closing a switch connecting a battery to the input terminals.

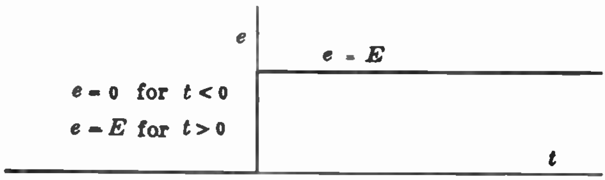


FIG. 1-24. Driving force (input voltage) which, when applied to the low-pass filter whose characteristics are given in Fig. 1-23, gives a response (output current) shown in Fig. 1-25.

By consulting a table of integrals of the type[10] (1-106)

$$F(\omega) = \frac{E}{2\pi \cdot j\omega} \tag{1-110}$$

From the (steady-state) ideal characteristic of the filter (Fig. 1-23), it is seen that $\underline{Y}(\omega)$ can be expressed

$$\left.\begin{aligned} \underline{Y}(\omega) &= 0 \quad \text{for } |\omega| > |\omega_1| \\ \underline{Y}(\omega) &= Y_0 \angle\theta \quad \text{for } |\omega| < |\omega_1| \end{aligned}\right\} \tag{1-111}$$

and since $\theta = -\omega t_d$, $\underline{Y}(\omega) = Y_0 \angle -\omega t_d = Y_0\, e^{-j\omega t_d}$

Hence the output current is

$$i(t) = \frac{Y_0 E}{2\pi} \int_{-\omega_1}^{\omega_1} \frac{e^{j\omega t} e^{j\theta}}{j\omega}\, d\omega \tag{1-112}$$

$$= \frac{Y_0 E}{2\pi} \int_0^{\omega_1} \frac{e^{j\omega(t - t_d)}}{j\omega}\, d\omega + \frac{Y_0 E}{2\pi} \int_{-\omega_1}^{0} \frac{e^{j\omega(t - t_d)}}{j\omega}\, d\omega$$

$$= \frac{Y_0 E}{\pi} \int_0^{\omega_1} \frac{\sin \omega\, (t - t_d)}{\omega}\, d\omega$$

$$= \frac{Y_0 E}{\pi} \int_0^{\omega_1 (t - t_d)} \frac{\sin y}{y}\, dy \tag{1-113}$$

9. See Sec. 1-11, subsection "Ideal Admittance Characteristics."

10. Campbell, George A., and Foster, Ronald M., Fourier Integrals for Practical Applications, Bell System Technical Publication B-584. This volume contains an extensive table of the values of integrals of the form (1-105) and (1-106), and may be used in much the same way an ordinary table of integrals is used.

where the limits of the integral in (1-112) have been made $\pm\omega_1$ since Y is zero outside this range; hence the integrand is zero outside $\pm\omega_1$. The integral (1-113) is one which cannot be integrated, but which has been tabulated for various values of the upper limit.[11] The output is shown in general form in Fig. 1-25.

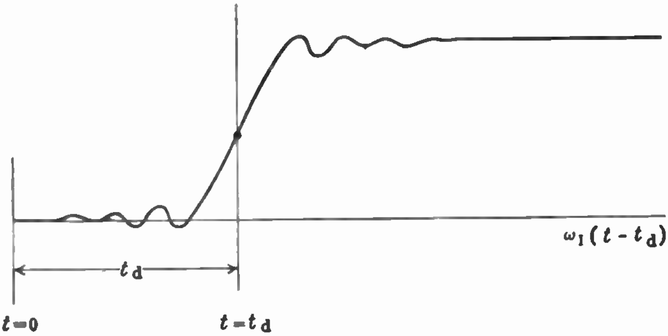


FIG. 1-25. Response of low-pass filter with admittance characteristics given by Fig. 1-23 to suddenly applied driving force given by Fig. 1-24.

Some other results of the use of the Fourier Integral in determining responses of a low-pass filter (ideal from the steady-state viewpoint) are shown in Fig. 1-26 and in Fig. 1-27. These have been taken from Sullivan.[12]

Although the above illustration has been devoted to a low-pass filter, this has been merely an example. Equation (1-108) is a powerful tool, so long as the transfer admittance characteristic is known. The low-pass filter case is cited by most writers because of the ease with which it can be handled, but eq. (1-108) can be applied rather generally, and if the mathematical process of integration becomes too difficult, numerical integration can be used to produce the result.

1-23. Response of Circuits to Non-sinusoidal Waves.- It is often desired to determine the characteristics of a circuit, for example an amplifier, over its entire working frequency range. As often done point by point, this is a rather tedious procedure, and if various circuit modifications are being tried, considerable time is consumed in determining the overall effect

11. Tables of Sine, Cosine and Exponential Integrals (2 vols.) available from the National Bureau of Standards, Washington, D.C.
12. Sullivan, W.L., Analysis of Systems with Known Transmission-Frequency Characteristics by Fourier Integrals, Electrical Engineering, Vol. 61, No. 5, May 1942.

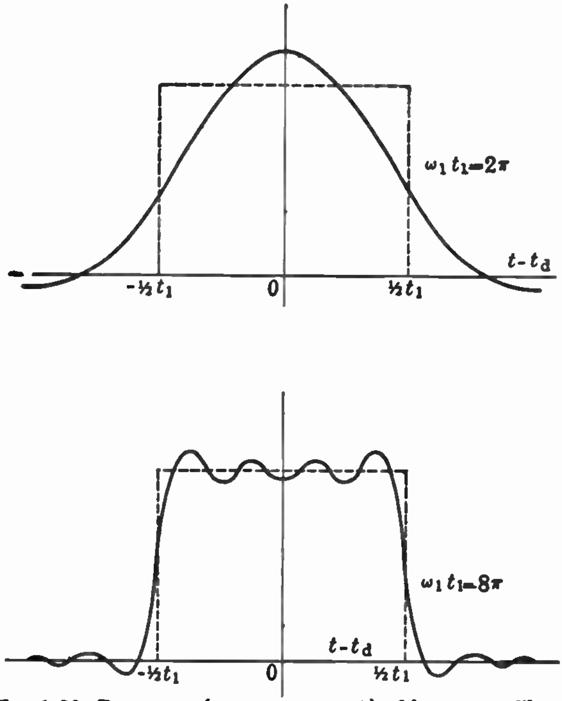


FIG. 1-26. Response (current output) of low-pass filter having admittance characteristics given by Fig. 1-28. Driving force (input voltage) is the rectangular pulse shown dotted.

of each modification.

It has been seen that pulses contain all frequencies, and rectangular periodic waves contain many multiple frequencies and hence offer the possibility, when applied to a circuit to be tested, of indicating at once practically the entire frequency characteristic--at least to the extent to which the user is capable of interpreting the results. It is desirable to develop a judgment of the meaning of response to square-wave testing by comparing response vs frequency curves and square-wave-test output curves for a variety of circuits, some with poor low-frequency response, some with poor high-frequency response, and so on. The output is ordinarily viewed on the screen of a cathode-ray oscillograph.

If the generator indicated in the circuit of Fig. 1-28 generates a square wave of adjustable fundamental frequency, then the variation of the resulting voltage across the condenser may have a variety of shapes depending on the frequency. If the ratio of condenser reactance at fundamental frequency to

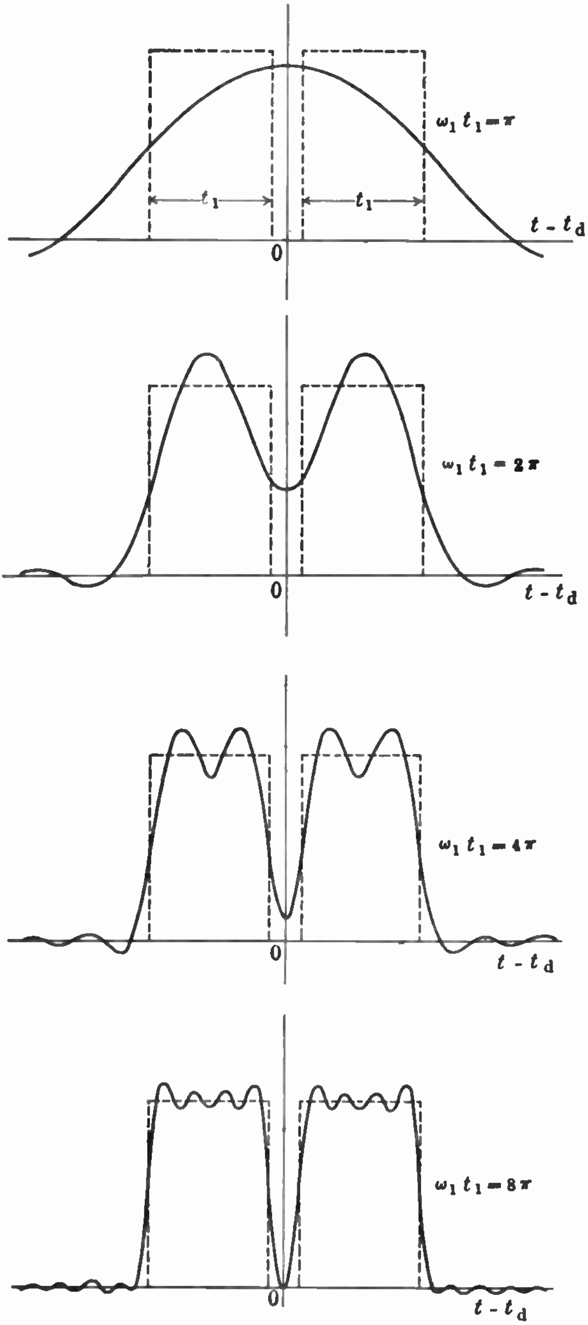


FIG. 1-27. Response of low-pass filter whose characteristics are given in Fig. 1-23 to a double, rectangular, pulse. Shape of input pulse is shown dotted on each figure; solid curve in each case is output current.

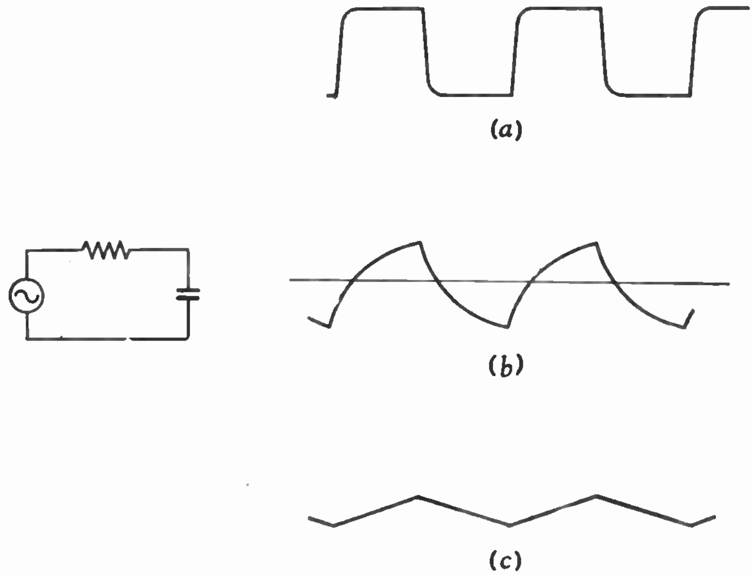


Fig. 1-28. R-C circuit with square-wave generator, and condenser voltage variations for three different applied frequencies. (General Radio Experimenter)

circuit resistance is 10 to 1, then the curve (a) represents the condenser voltage variation. If the ratio is 1 to 1, curve (b) results, and (c) represents the conditions for a ratio of 0.1 to 1.

In Fig. 1-29 is shown the amplitude response for a single-stage amplifier, first as a function of frequency, and in (a), (b), and (c) respectively the square-wave response at frequencies of 1, 10, and 100 cps respectively. Although the amplitude response curve is flat at 100 cps, the response to a 100-cps square wave is not square, owing to phase shift differing for the different harmonics.

A number of tests on transformers of different qualities, in which the frequency characteristics are compared with the output to square-wave excitation at 60 and 2000 cycles per second, are illustrated in Fig. 1-30.

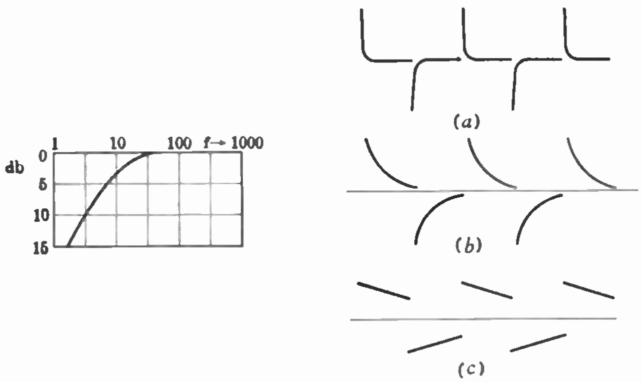


FIG. 1-29. Frequency response of a single-stage amplifier, and square-wave response at three applied frequencies. (General Radio Experimenter.)

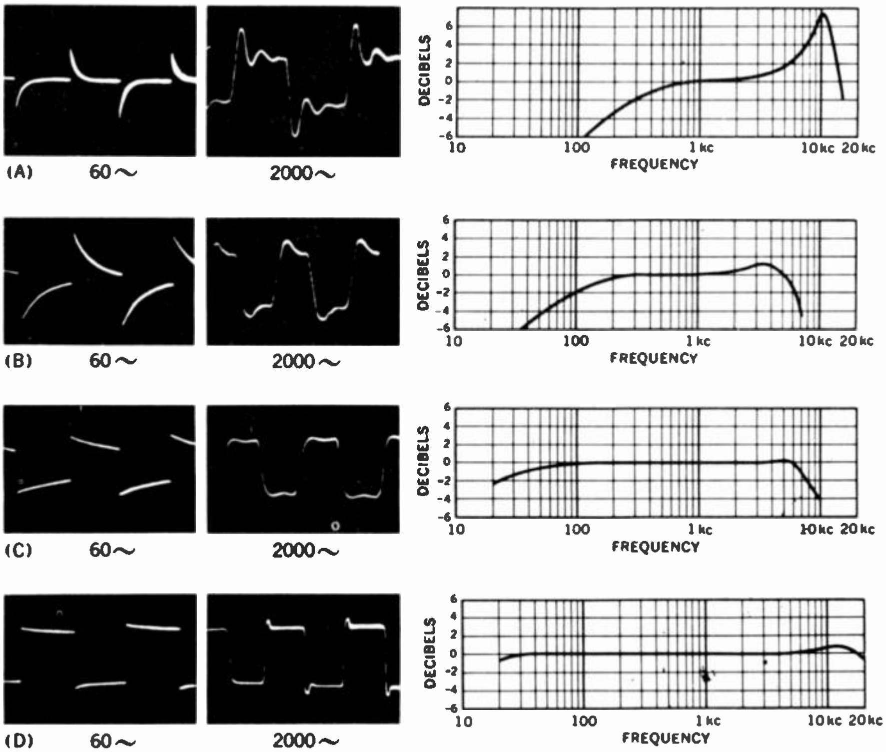


FIG. 1-30. Square-wave response (left) and amplitude response characteristic (right) characteristics for four different audio-frequency transformers. (The amplitude response characteristics are for sine-wave impressed voltages; the dates refer to construction of the transformers. The curves show the improvement achieved; the square-wave tests do likewise. General Radio Experimenter.)

Chapter 2

FUNDAMENTALS OF TUBES; POWER SUPPLIES

Because of the large amount of advanced material that must be covered in a course in ultra-high-frequency techniques, it is necessary to assume that the student is already familiar with the simpler aspects of the theory of electron tubes and circuits and with the physical phenomena upon which the operation of electron tubes is based. The brief summary contained in the paragraphs that follow may be of value in suggesting topics that should be reviewed by the student.

2-1. Electrons and Ions.- The operation of electron tubes of all types is dependent upon the presence and movement of electrons within tubes and in electrical conductors connecting the electrodes. In tubes containing gas or vapor the operation also depends upon the presence and movement of positive ions, the most common type of which is the positively charged particle that results from the removal of an electron from a neutral gas atom or molecule when the atom or molecule is struck by a rapidly moving electron. In order that an electron shall be able to ionize an atom or molecule it must possess kinetic energy at least equal to that which it would acquire in being accelerated from rest by the electric field existing between two electrodes having a potential difference equal to the ionization potential of the gas. An appreciable time is required for a gas to become ionized, and a very much longer time (100 to 1000 μsec) for the ionized particles to be removed from the interelectrode space.

A group of free charges in space is called space charge. If the charge is of one sign only, or if the density of charge of one sign in a given volume exceeds that of the other sign, an electric field is set up by the space charge.

2-2. Electron Emission.- The electrons that make possible the flow of current between the electrodes of an electron tube are ordinarily removed from the electrodes by thermionic emission, secondary emission, photoelectric emission, or field emission. The phenomenon of thermionic emission is in many respects similar to the vaporization of a liquid or solid. Metals contain large numbers of electrons which are loosely bound to the atoms and are therefore able to move about within the metals. As the result of collisions with atoms and molecules these free electrons acquire random velocities, the average value of which increases with the temperature of the metal. In passing through the surface of the metal an electron loses an amount of kinetic energy, called the electron affinity, that depends upon the kind of metal and the condition of its surface. Electrons adjacent to the surface, and having a component of velocity toward the surface of such magnitude that the associated kinetic energy is at least equal to the electron affinity, pass through the surface of the metal and into the space beyond. If the space outside the metal does not contain an applied accelerating electric field, or if the field is small, a space charge forms outside the metal and produces a repelling field which causes most of the electrons to re-enter the metal. The electrons that move away from the emitter produce an emission current. The emission current is increased by an accelerating field produced at the surface of the metal by means of a positively charged conductor near the surface of the emitter, called the anode or plate. The emission current increases with anode voltage toward the limiting value obtained when all the electrons are drawn away as soon as they are emitted. This saturation current is given by Richardson's equation:

$$I_s = AT^2 e^{-w/kT} \qquad (2\text{-}1)$$

in which I_s is the saturation current per unit area of the emitter (amperes per unit area), T is the absolute temperature (degrees C.), w is the electron affinity of the emitter (ergs), k is a universal constant (= 1.37×10^{-16} erg per degree C), and A is a constant. Richardson's equation shows the necessity of using metals having low electron affinity in order to obtain high emission at relatively low temperature.

Secondary emission is emission of electrons from a surface as the result of bombardment of the surface by other electrons or by positive ions. Appreciable secondary emission may be obtained when the surface is bombarded by electrons which have been accelerated through a potential difference of only a few volts. The amount of secondary emission depends upon the substance from which emission takes place, upon the condition of the surface, upon the temperature of the emitter, and it increases with velocity of the bombarding particles. It is considerably larger for bombardment by electrons than for bombardment by positive ions. A single primary electron may eject as many as 8 or 10 secondary electrons.

Field emission, i.e., emission of electrons solely as the result of the existence of an electrostatic field at the surface of a metal, takes place only at extremely high field strengths and is of importance principally in arc discharges.

Photoelectric emission, i.e., the emission of electrons as the result of electromagnetic radiation falling upon the surface of a metal, is of importance in glow and arc discharges and in phototubes, none of which will be taken up in this text.

2-3. Practical Thermionic Emitters.- In order to obtain high emission currents for a given amount of heating power, and in order to prevent vaporization of the emitter, it is necessary to use emitting materials that have a low electron affinity. Although both pure tungsten and tantalum are used in high-voltage transmitting tubes, most electron tubes make use of thoriated tungsten or oxide-coated emitters which have much lower electron affinity. Before either a thoriated tungsten or an oxide-coated cathode will give appreciable emission at normal operating temperatures it must be "activated" in a manner discussed in practically all books dealing with the theory and applications of electron tubes. In the filamentary type of emitter, emission takes place directly from the surface of the conductor through which the heating current passes. In the heater type of emitter, the emission takes place from an oxide-coated surface, usually in the form of a cylindrical nickel sleeve, which is heated by means of an electric heating coil electrically insulated from the emitting surface. Although filamentary emitters heat more rapidly than heater-type cathodes, the latter have the advantage that the cathode and the heater are not connected electrically; hence several tubes may be operated from the same source of heater current with their cathodes at different potentials. The alternating current of the heater does not generally influence the electrode currents.

2-4. Limitation of Current by Space Charge.- As pointed out in Sec. 2-2, the space charge that collects near the surface of an emitter produces a field at the surface which tends to cause the electrons to re-enter the surface, and a high positive voltage must be applied to an adjacent anode in order to draw electrons away as fast as they are emitted. At any voltage less than that necessary to produce saturation current, the field at the cathode is the resultant of the accelerating field caused by the applied positive anode voltage and the retarding field produced by the negative space charge. A theoretical expression, known as Child's law, for the anode current in terms of the anode voltage under the assumption that electrons have zero velocity upon emission from the cathode and that the electrodes are unipotential surfaces, may readily be derived by the application of electrodynamics. For plane parallel electrodes whose area S is large in comparison with the spacing d between them, Child's law is

$$i_b = 2.34 \times 10^{-6} \frac{S e_b^{3/2}}{d^2} \qquad (2\text{-}2)$$

in which i_b is the anode current in amperes and e_b is the anode voltage in volts, and S and d^2 must have the same units. For concentric cylindrical electrodes whose length h is large in comparison with the spacing between them, Child's law is

$$i_b = 14.68 \times 10^{-6} \frac{h e_b^{3/2}}{\beta r} \qquad (2\text{-}3)$$

in which r is the radius of the anode and β is a factor whose value depends upon the ratio of the radius of the anode to that of the cathode; β has the approximate value $\frac{1}{4}$ for a ratio 2, $\frac{1}{2}$ for a ratio 3, and 0.9 for a ratio 8. The fact that practical diodes do not satisfy the assumptions made in the derivation of eqs. (2-2) and (2-3) causes the characteristics of practical diodes to differ from the theoretical characteristics, but these equations do indicate the importance of close anode-cathode spacing if high anode current is desired for a given value of anode voltage e_b.

Because the plate of a diode does not ordinarily become hot enough to emit electrons, no current flows when the anode is made more than about 0.5 volt negative. This fact makes possible the use of diodes in modulation, detection, and in power rectification.

2-5. The Triode.- The introduction of a third electrode between the cathode and plate of a vacuum tube makes possible the variation of plate current by small variations of the voltage of this third electrode relative to the cathode. If this electrode is maintained negative relative to the cathode, no space current flows to or from it, and so at frequencies lower than a few hundred megacycles, the control is accomplished without the expenditure of energy in the control electrode circuit. In order that electrons may pass to the plate, the third electrode is ordinarily made in the form of a wire grid and is known as the grid. Because the grid is located between the cathode and the plate, many of the electric field lines from the plate terminate on the grid and only a portion terminate on the cathode. The grid, on the other hand, is not shielded from the cathode. For this reason and because the distance from cathode to grid is less than that from cathode to plate, the resultant field at the cathode is affected much more by variations of grid voltage than by equal variations of plate voltage, and the grid is much more effective than the plate in controlling the plate current. This fact may be expressed in the form of the functional equation:

$$i_b = F(e_b + \mu e_c + \varepsilon) \qquad (2\text{-}4)$$

in which μ, the amplification factor, is a quantity greater than unity, ε is a quantity that takes into account the initial velocities of the electrons after emission and the contact differences of potential existing between the electrodes as the result of differences of electron affinity of the metals of which the electrodes are made, and e_c is the grid voltage. The amplification factor depends upon grid wire diameter and spacing, upon electrode spacing, and upon electrode voltage and currents. Except in the case of tubes in which the grid wires are not evenly spaced, the variation of μ with electrode voltages is sufficiently small so that it may often be assumed to be constant over the working range of current and voltage. The amplification factor μ may be defined mathematically by

$$\mu = -\frac{\partial e_b}{\partial e_c} \qquad (i_b = \text{const.}) \tag{2-5}$$

and μ is thus seen to be the negative of the slope of the curve of e_b vs e_c for constant plate current. It may be found approximately by taking the ratio of a small increment of plate voltage to the corresponding increment of grid voltage at constant plate current in the vicinity of the voltages for which the value is desired. These increments may be found from the family of curves of i_b vs e_b at constant e_c.

Two other tube factors are very important in the analysis of vacuum tube circuits. These are the a-c plate resistance,[1] which is defined mathematically by

$$r_p = \frac{1}{\partial i_b/\partial e_b} \tag{2-6}$$

and the (grid-plate) transconductance, which is defined mathematically by

$$g_m = \partial i_b/\partial e_c \tag{2-7}$$

From eq. (2-6) it is seen that r_p is the reciprocal of the slope of the plate characteristic, i_b vs e_b, for the given operating voltages, and (2-7) shows that g_m is the slope of the transfer characteristic, i_b vs e_c at the given voltages. The plate resistance corresponding to given values of electrode voltage may be found most readily for a triode by drawing a tangent to the i_b - e_b curve at the point in question and computing the value of the reciprocal of the slope of the tangent in volts per ampere, or ohms. Differentiation of eq. (2-4) shows that $g_m = \mu/r_p$. The value of the transconductance corresponding to given values of electrode voltage may therefore be found from the ratio of the amplification factor to the plate resistance. It may also be found from the transfer characteristic, but transfer characteristics are not often available. It should be emphasized that the transconductance and plate resistance both vary greatly with electrode voltages, and that the value listed for one set of voltages cannot, therefore, necessarily be used at other voltages.

1. The d-c plate resistance, which is seldom used, is e_b/i_b.

Typical triode characteristics[2] are those of the 6F6 tube with triode connection, or of the 6C5.

2-6. Tetrodes and Pentodes.- In certain applications of electron tubes it is important to reduce to a low value the capacitance between the plate and the control grid. This can be accomplished by introducing a second grid, called the screen grid, between the control grid and the anode. In order to pull the electrons away from the cathode and at the same time to allow a large portion of them to be drawn to the plate, the screen grid is operated at a voltage less than the steady component of anode voltage. Because of the fact that the screen grid also shields the plate from the cathode, that is, allows few field lines from the plate to terminate upon the cathode, the introduction of the screen grid decreases the effect of the anode voltage on electrons at the cathode, and thus increases the plate resistance and the amplification factor of the tube. Since the screen voltage is maintained constant in the operation of the tube, whereas the plate voltage may vary, the voltage of the plate may fall below that of the screen. Secondary electrons emitted from the plate are then drawn to the screen and thus reduce the plate current. This causes the i_b-e_b curves to have negative slope over a range of plate voltage, as shown by the characteristics of the 24A tetrode, which is a typical tetrode. This is undesirable in the use of tubes in voltage and power amplification, as it restricts the range over which the plate voltage may vary without introducing excessive distortion. The difficulty may be eliminated by the introduction between the screen and the plate of a third grid, called the suppressor grid, maintained at cathode potential. Secondary electrons are then returned to the plate by the field between the suppressor grid and the plate. The additional shielding of the suppressor grid causes the plate resistance and amplification factor of pentodes to be even higher than those of tetrodes. The objectionable effects of secondary emission from the plate can also be prevented by concentrating the cathode-plate space current into beams so as to produce a high-density space charge near the plate. The field due to the negative space charge returns secondary electrons to the plate in a beam power tube in the same manner as the field between suppressor grid and plate does in the suppressor pentode. Because the field between the screen and the plate is more uniform in a beam power tube than when a suppressor grid is used, the characteristic curves of beam power tubes are of a

2. For these and other characteristics, consult a tube manufacturer's handbook such as that of the Radiotron Division, RCA Manufacturing Co., Harrison, N. J. The student should become thoroughly acquainted with tube manuals of this type.

more desirable form than those of suppressor pentodes. The 6SJ7 and the 6F6 tubes illustrate typical suppressor-pentode characteristics, and the 6L6 typical beam-pentode characteristics.

In certain applications of tubes it is desirable that the transfer characteristic, i_b vs e_c, should approach the voltage axis sharply. This result is attained by the use of a grid having uniform spacing between wires of constant diameter. Such a grid is called a sharp-cutoff grid. In other applications, on the other hand, it is essential that the transfer characteristic should approach the voltage axis gradually. Such a characteristic results when the grid-wire spacing or the diameter of the grid wires varies along the length of the grid. Such a grid is called a variable-mu grid, or a remote-cutoff grid.

Because the electrode currents of vacuum tubes are not linearly related to the electrode voltages, the application of alternating voltages to the electrodes, in addition to steady voltages, results in changes of average as well as instantaneous current. If the external circuits contain resistances, then the average electrode voltages also vary. The steady values of electrode voltages and currents existing when no alternating voltages are applied to the electrodes are called the static operating (quiescent) voltages and currents. The average values assumed when alternating voltage excitation is applied are called the dynamic operating values. The average value of grid voltage is called the grid bias. In order to prevent the flow of grid current, the control-grid bias is usually made negative with respect to the cathode. The crest value of alternating grid voltage is called the grid swing.

2-7. Transit Time.- At frequencies lower than a few hundred megacycles, the time taken for electrons to traverse the interelectrode spaces is so short, in comparison with the period, that the plate current may be assumed to respond instantaneously to changes of electrode voltages. As will be seen, however, at ultra-high frequencies the transit time of electrons is an extremely important factor in the behavior of vacuum tubes, and in some tubes the period, say of a 1000-megacycle current, may be approximately equal to the transit time. The effect of transit time on high-frequency oscillators is discussed in Ch. 10.

2-8. Series Expansion for Electrode Currents.- Because the functional eq. (2-4) is complicated in form, and cannot be put into simple explicit form, this equation is of little value in the analysis of vacuum-tube circuits. The alternating component of plate current may, however, be expressed in the form of the series

$$i_p = a_1 e + a_2 e^2 + a_3 e^3 + \ldots\ldots \tag{2-8}$$

in which

$$e = e_g + v_p/\mu \qquad (2\text{-}9)$$

where e_g is the alternating voltage between the grid and the cathode, and v_p is the impressed alternating voltage in the plate circuit. The coefficients a_n are constants involving the external circuit parameters, the plate resistance r_p, the amplification factor μ, and their derivatives evaluated at the operating point.

If the external circuit contains inductance and capacitance, the series (2-8) must be replaced by another of the same type[3] in which each coefficient is a complex number whose modulus value is to multiply e, e^2, etc., and whose phase angle is to be added to the various sine terms contained in e, e^2, etc. The first coefficient is

$$\underline{a}_1 = \frac{\mu}{r_p + \underline{Z}_o} \qquad (2\text{-}10)$$

where $\underline{Z}_o$ is the impedance in the external plate circuit.

Because of the complicated form of the higher order coefficients, the series expansion for plate current is ordinarily not of great value in the numerical solution of particular circuits. It is, however, of importance in the general analysis of amplification, detection, and modulation. One application is in the proof that the non-linearity of the vacuum tubes results in the generation by the tubes of harmonic and intermodulation frequencies not present in the exciting voltages impressed upon the circuits. The proof consists merely in substituting in eq. (2-8) a voltage e which is sinusoidal in form or is the sum of two or more sinusoidal voltages. The series expansion may also be used to show that the introduction of impedance into the plate circuit tends to straighten the transfer characteristic and thus to reduce the generation of intermodulation and harmonic frequencies.

Equations (2-8) to (2-10), with suitable changes in symbols, may also be applied to electrodes other than the plate.

2-9. The Equivalent Plate Circuit.- In the solution of many vacuum tube problems, particularly when the excitation voltage is small and the load impedance high, sufficient accuracy is obtained by using only the first term of the series expansion. This is equivalent to assuming that the tube is a linear circuit element. Examination of the first term of the series expansion for plate current leads to the equivalent-plate-circuit theorem for amplification, which states that the tube may be replaced by assuming there is connected between the cathode and the plate a

---

3. The derivation of these series is given in detail in McIlwain and Brainerd, "High Frequency Alternating Currents," Wiley, New York, 1939.

fictitious generator of voltage $v_p + \mu e_g$ in series with the a-c plate resistance r_p. Once the equivalent circuit[4] has been formed, it may be solved just as any other a-c circuit, and no consideration need be given the fact that the actual circuit contains a tube.

Use of the following procedure insures that voltage and current polarities are correct in the equivalent circuit:

1. In the actual circuit diagram, show the instantaneous grid excitation voltage v_g (the alternating voltage impressed in the grid circuit) in such polarity as to make the grid positive.

2. Show the instantaneous plate current i_p flowing into the plate.

3. Insert the equivalent voltage μe_g, in series with the operating plate resistance r_p, between the plate and the cathode, choosing the polarity of the equivalent voltage so that it would cause i_p to flow in the direction indicated in 2.

4. Assume directions for the other instantaneous circuit currents.

5. Delete (or show dotted) the tube symbol, the batteries, and all circuit elements not coupled to the plate (such as the screen circuit).

6. Redraw the resulting equivalent circuit in the form in which it may be most readily analyzed.

The value of the instantaneous grid voltage e_g may differ from the exciting voltage v_g applied from an external source if the plate circuit is coupled in any manner to the grid circuit, as, for instance, in Fig. 2-1. In general, therefore, it is necessary to evaluate e_g in terms of v_g and circuit parameters and currents before the circuit can be solved. The instantaneous voltage e_g is usually most readily found from the actual circuit and has a complex equivalent equal to the vector sum of all alternating voltages between the cathode and the grid along any continuous path.

With suitable changes of symbols, the above procedure may be used for the circuits of other electrodes than the plate. The method of forming an equivalent plate circuit is illustrated in Fig. 2-1. Summation of alternating voltages between the

---

4. Certain precautions must be taken in the application of the equivalent-plate-circuit theorem. Since grid current flows only when the grid is positive, the flow of grid current through external impedances may result in non-sinusoidal grid voltages. Hence the theorem is ordinarily applicable only when the grid is negative. In setting up the equivalent circuit the interelectrode capacitances must be drawn external to the tube. Because the equivalent circuit applies only to the fundamental components of currents, it is of little use in the analysis of detection and modulation, although similar circuits, based upon the higher terms of the series expansion, are sometimes useful. The equivalent circuit is not applicable at frequencies so high that electron transit time is important. The theorem is used particularly in connection with amplifiers and oscillators. Chapter 3 (Amplification) is based to a large extent on results obtained by replacing a tube by the fictitious generator in series with r_p between cathode and plate, and solving the resulting circuits.

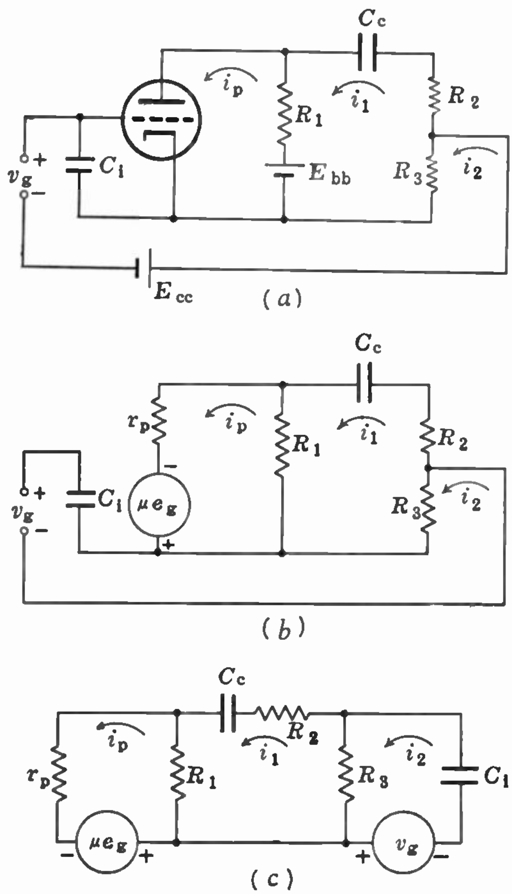


FIG. 2-1. (*a*) Tube circuit. (*b*) Corresponding equivalent plate circuit. (*c*) Equivalent plate circuit after rearrangement.

cathode and the grid shows that, writing e_g, v_g, i_1 and i_2 in complex form, $\underline{E}_g$ is equal to $-\underline{I}_2/j\omega C_1$ or to $\underline{V}_g - (\underline{I}_2 - \underline{I}_1)R_3$.

2-10. Electrode Capacitances.- The operation of electron tubes may be greatly affected by the capacitances between electrodes (see Fig. 3-7). The effective input capacitance may be much greater than the physical interelectrode capacitances. It is shown in Sec. 3-3 that the effective input capacitance between the grid and cathode of a vacuum tube at frequencies below a few hundred megacycles is approximately equal to

$$C_i = C_{gk} + (1 + A)C_{gp} \qquad (2\text{-}11)$$

in which C_{gk} is the grid-cathode capacitance, C_{gp} is the grid-plate capacitance, and A is the magnitude of the voltage amplification of the tube and circuit. Since A may be large, it is evident that the effective input capacitance may be large, even though the interelectrode capacitances are small. Since A cannot exceed μ, the effective input capacitance cannot exceed $C_{gk} + (1 + \mu)C_{gp}$.

2-11. Plate Diagram. A very useful tool in the analysis of many vacuum tube problems is the plate diagram, which consists of the family of i_b-e_b characteristics on which is indicated the path of operation, or locus of simultaneous values of current and voltage assumed when the tube is excited. In general, the path of operation is approximately elliptical in form,[5] and can be constructed only by methods of successive approximation. Much useful information can be obtained, however, if it is assumed that the load is non-reactive. The path of operation, or load line, is then linear, and can be readily constructed. A typical plate diagram for a triode, constructed under the assumption that the average plate current does not change appreciably when the tube is excited, is shown in Fig. 2-2.

In Fig. 2-2 the static load line is the locus of all corresponding values of steady voltage and current that can be assumed with the given load and plate supply voltage E_{bb} when no alternating voltage is impressed upon the electrodes. This line

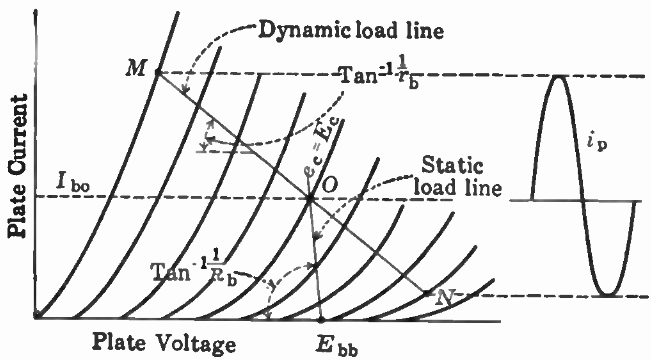


Fig. 2-2. Simplified plate diagram for a circuit in which the a-c resistance of the load exceeds the d-c resistance.

5. Because if $i = I \cos \omega t$ and $e = E \cos (\omega t + \theta)$ where I, E, ω, and θ are constants, then i vs e is an ellipse, as can be shown by eliminating t and getting i in terms of e. Ellipses such as this can be readily produced on a cathode-ray oscillograph. In the case of the vacuum tube the locus is not an ellipse because the varying current does not vary as the voltage, but as indicated by (2-8).

passes through a point on the voltage axis corresponding to the plate supply voltage, and has a negative slope equal to the reciprocal of the d-c resistance of the plate circuit, expressed in amperes per volt. The intersection of this line with the plate characteristic corresponding to the grid bias, E_c, determines the static operating point O. The dynamic operating line is the locus of all simultaneous values of current and voltage assumed during the cycle when alternating voltage is impressed upon the grid. This line passes through the operating point and has a negative slope, expressed in amperes per volt, equal to the reciprocal of the a-c resistance of the plate load.

From the intersections of the dynamic load line with the static plate characteristics it is possible to find corresponding values of plate current and grid voltage from which the dynamic transfer characteristic of the tube and load may be constructed. This may in turn be used to construct a wave of plate current corresponding to a given wave of grid voltage. From the intersections of the dynamic load line with the plate characteristics it is also possible to find the change of plate voltage corresponding to a given change of grid voltage. Since the change of plate voltage is equal in magnitude to the change of voltage across the load, the ratio of the change in plate voltage to the change in grid voltage, corresponding to points on the dynamic load line, gives the approximate voltage amplification. The power output may be determined by finding one-eighth the area of the rectangle, sides parallel to the axes, which has as a diagonal the ends of the dynamic load line for the given grid swing. The application of simple formulas for harmonic content also makes it possible to determine the harmonic content of the curve of plate current, and thus the percentage distortion.

2-12. Power Supplies.- Practically all vacuum tubes are dependent in their operation upon a source of high direct voltage. Ordinarily an electron-tube rectifier constitutes the most satisfactory source of high-voltage power for this purpose. Among the advantages of electronic power supplies over rotating machines are their relatively low cost, their compactness and light weight, their quietness of operation, effectiveness of control of voltage, and the ease with which they may be turned on and off. Since the output voltage of a rectifier is pulsating, and the voltage used in most tube circuits must be nearly free of pulsation, rectifiers that supply direct voltages to tubes must usually be followed by smoothing filters.

The alternating component of unidirectional voltage from a rectifier or generator used as a source of direct voltage is called the <u>ripple voltage</u>. The ripple voltage is not generally sinusoidal and therefore possesses both fundamental and harmonic components. The effectiveness of a smoothing filter increases with the frequency of the voltage input to the filter and the amplitudes of the ripple harmonics decrease with the order of the harmonics. A filter which is designed to produce adequate filtering at the fundamental ripple frequency will, therefore, usually reduce ripple harmonic voltages to negligible values. Hence the degree to which ripple voltage is objectionable can best be specified by the ratio of the amplitude of the fundamental component of the ripple voltage to the average (direct) value of total voltage. This ratio is known as the ripple factor.

The ripple that can be tolerated at the load depends upon the purpose for which the power supply is to be used. In the microphone circuit of a radio transmitter the ripple factor should not exceed 0.005 per cent. In audio-frequency amplifiers, ripple factors may lie in the range from 0.01 to 0.1 per cent, and even 1 per cent may sometimes be used without objectionable hum resulting. In voltages for cathode-ray oscillograph tubes the ripple should be less than 1 per cent. In certain types of electronic control devices, such as those making use of phototubes to operate relays, the allowable ripple may sometimes be so high that no smoothing filter need be used. The effectiveness of a smoothing filter in removing a component of ripple voltage of any frequency is indicated by the effect the filter would have upon a sinusoidal voltage of that frequency. The ratio of the amplitude of a sinusoidal voltage of given frequency impressed upon the input of a filter to the amplitude of the resulting sinusoidal output voltage will be termed the smoothing factor at that frequency. The following symbols will be used:

Supply frequency..........................	f
Fundamental ripple frequency..............	f_r
Amplitude of fundamental component of ripple voltage at filter input..............	E_{r1}
Average value of total output voltage......	E_{dc}
Ripple factor, E_{r1}/E_{dc}, at input to filter.	ρ
Smoothing factor at any frequency..........	α
Smoothing factor at fundamental ripple frequency..................................	α_1

Primed symbols will be used to indicate ripple voltage and its components at the output of the filter (across the load).

2-13. Rectifier Circuits.- Figure 2-3 shows the circuits of the two rectifiers used most commonly to supply voltages for the operation of equipment discussed in later chapters of this book, the single-phase half-wave rectifier, and the single-phase full-wave rectifier. Curves of positive secondary transformer voltage and of load current for resistance and inductive loads are shown in Fig. 2-4. These curves are obtained under the assumption that tube and transformer voltage drops are negligible.

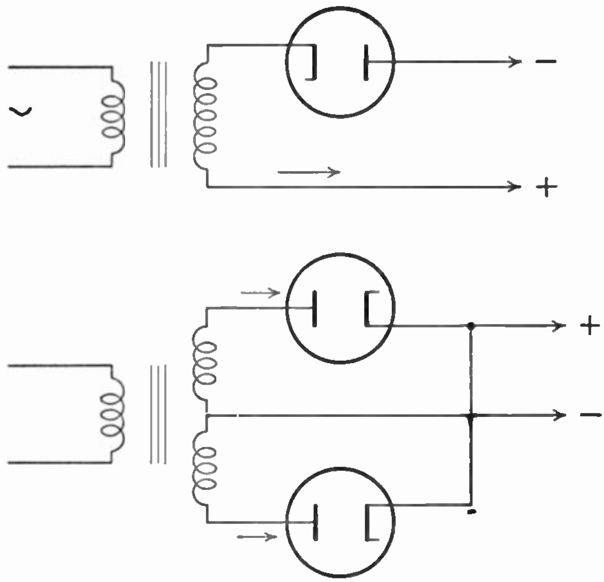


FIG. 2-3. Single-phase half-wave and full-wave rectifiers.

When the load (including the filter) is non-reactive, current flows in each anode only during the portion of the cycle in which the secondary voltage applied to that anode is positive and the instantaneous current is proportional to the instantaneous secondary voltage. The current waves are therefore of the form shown by the dotted curves. Inductance in the load tends to prevent the current from building up or dying down, and so changes the current wave forms to those shown by the solid curves. If the inductance can be made large enough, the load current becomes nearly constant. In the full-wave circuit the current into each anode is then nearly a rectangular pulse.

Important information concerning half-wave and full-wave single-phase rectifier circuits is listed in Table 2-I. In

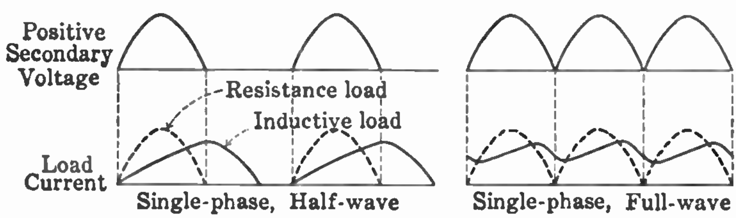


FIG. 2-4. Curves of positive secondary voltage and load current for single-phase rectifiers.

deriving the values listed in this table, it is assumed that tube and transformer drops are negligible and that the load is resistive or inductive. Capacitive load will be discussed in a separate section. The chief merits of the single-phase half-wave

Table 2-I

Type of circuit	Single-phase half-wave: Inductance in series with load	Single-phase half-wave: Resistance load, without choke	Single-phase full-wave Inductance in series with load
Secondary volts per leg ÷ direct voltage..	2.22	2.22	1.11
Primary volts per leg ÷ direct voltage....	2.22	2.22	1.11
Primary current per leg ÷ direct current..	0.707	1.57	1.0
Secondary current per leg ÷ direct current	0.707	1.57	0.707
Primary kva ÷ d-c watts..................	1.57	3.49	1.57
Secondary kva ÷ d-c watts................	1.57	3.49	1.11
Peak inverse tube voltage ÷ direct voltage	3.14	3.14	3.14
Average tube current ÷ direct current.....	0.707	1.57	0.707
Peak tube current ÷ direct current........	1.00	π	1.00
Fundamental ripple frequency f_r in terms of supply frequency f..................	f	f	2f
Ripple factor ρ..........................	13*	1.57	0.667
Second-harmonic ripple factor ρ_2..........	1.77*	0.667	0.133
Third-harmonic ripple factor ρ_3...........	1.54*	0	0.057

*Values change with ratio of load resistance to series inductance.

rectifier are its simplicity and somewhat lower cost. Another advantage is that when the output of the rectifier is shunted by a condenser, approximately twice the output voltage can be obtained at light loads for a given number of total secondary turns with the half-wave circuit as with the full-wave circuit. These advantages are offset, however, by higher ripple amplitude and lower ripple frequency, which may necessitate the use of a more expensive and bulkier filter; by poorer voltage regulation; and by low transformer efficiency. The low transformer efficiency is caused by the facts that anode current flows during only one-half of the cycle and that rectified direct current flows through the secondary and saturates the core. The half-wave circuit is used principally in applications in which the load current is small, as in supplying the anode voltages for cathode-ray tubes, or grid bias for amplifiers.

2-14. Condenser Filter.- The simplest type of smoothing filter consists simply of a condenser shunting the output of the rectifier, as shown in Fig. 2-5a. Anode current flows only when the induced secondary voltage of half the transformer exceeds the condenser voltage. Since the charging current of the condenser

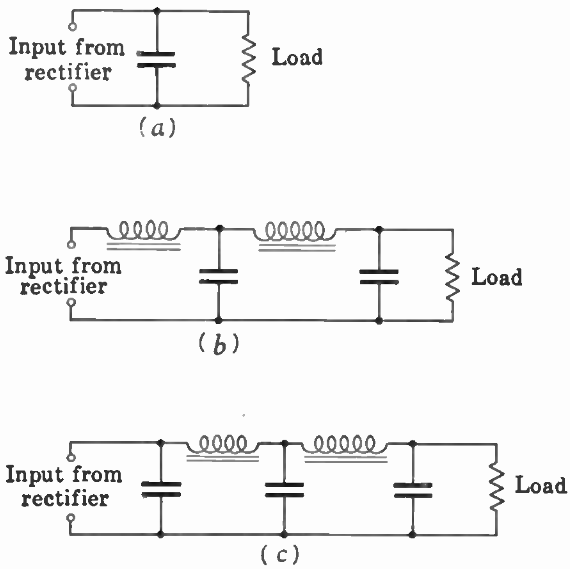


FIG. 2-5. Three types of smoothing filters: (*a*) condenser; (*b*) choke condenser (L-section); (*c*) choke-condenser with condenser input (pi-section).

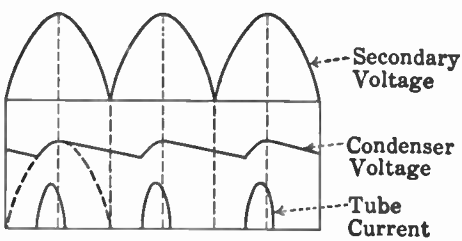


FIG. 2-6. Approximate wave form of condenser voltage and tube current in full-wave single-phase rectifier with resistance-capacitance load.

is limited only by the reactance of the transformer and the tube voltage drop, the condenser voltage rises nearly as fast as the induced secondary voltage until induced voltage reaches its maximum value. Unless the load current is very large, the induced voltage then decreases more rapidly than the condenser voltage, and so anode current cannot flow. The condenser discharges through the load during the remainder of the cycle, the condenser voltage falling exponentially at a rate determined by the condenser capacitance and the load current. Figure 2-6 shows the wave form of the condenser voltage and of the anode current. At light loads the ripple factor is given approximately by the relation

$$\rho = \frac{1}{\pi f_r RC} \qquad (2\text{-}12)$$

in which R is the resistance of the load, and C is the filter capacitance. Equation (2-12) holds for both the full-wave and the half-wave circuit. Since f_r is only half as great in the half-wave circuit as in the full-wave circuit, the ripple is twice as great in the half-wave circuit. At light load the direct voltage output of the half-wave circuit approximates the crest secondary voltage and that of the full-wave circuit approximates the crest voltage of half the secondary. Because the rate at which the condenser discharges increases with increase of load current, the average voltage across the load falls rapidly with increase of load current. For this reason the simple condenser filter results in poor voltage regulation and is used only in applications in which the load current is small.

2-15. Voltage Doubler.- Figure 2-7 shows a form of single-phase half-wave rectifier which gives a direct voltage that approximates twice the crest secondary voltage at light loads. The two condensers, which are charged in alternate halves of the cycle through two rectifier tubes, are connected in series so

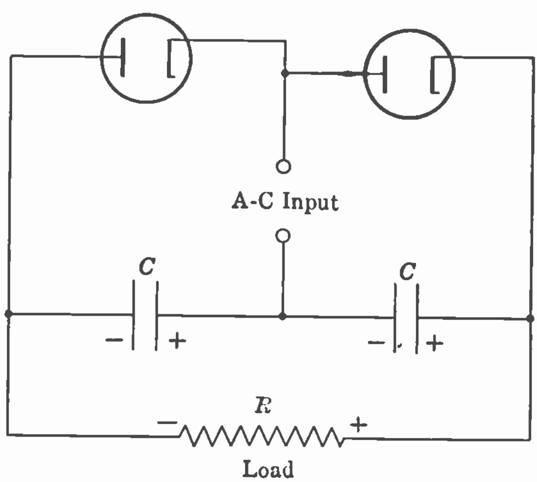


FIG. 2-7. Voltage doubler.

that their voltages add. The direct load voltage is twice the average voltage of each condenser, and the variation of load voltage (ripple) is equal to the variation of voltage of either condenser. The ripple frequency of the load is twice that of each condenser, or twice the supply frequency. The ripple factor at light loads approximates the value

$$\rho = \frac{1}{\pi f R C} \tag{2-13}$$

in which f is the supply frequency, R is the resistance of the load, and C is the capacitance of each condenser.

One advantage of the voltage doubler is that it may sometimes be used to furnish a required high direct voltage without the use of a transformer between the a-c supply and the rectifier. The circuit requires the use of two filament transformers, however, or of a full-wave rectifier tube with two independent cathodes. The 25Z5 tube is a heater-type rectifier tube designed for this type of service. More complicated circuits may be used to obtain a direct voltage approximately equal to four times the crest alternating voltage.[6] Voltage doublers, like the simple condenser filter, may be followed by other forms of smoothing filters. Because of their poor voltage regulation and high peak tube current, the use of voltage doubler circuits is usually limited to applications in which the required current output is small.

2-16. Choke-Condenser (L-Section) Filters.- Figure 2-5b shows the type of filter that is used in most power supplies designed to furnish appreciable load current. The exact analysis of a multi-section filter of this type is rather involved. Because the impedance that shunts a condenser is large, however, in comparison with the reactance of the condenser if adequate filtering is to be obtained, the shunting impedance may be neglected for many practical purposes. The smoothing factor α of a single section consisting of a series inductance and a shunt condenser is equal to the alternating voltage impressed across the series combination of choke and condenser divided by the alternating voltage appearing across the condenser. Under the assumption that the impedance shunting the condenser is so high that it may be neglected, this ratio is approximately equal to the vector sum of the inductive and capacitive reactances divided by the capacitive reactance.

$$\alpha = \frac{X_L - X_c}{X_c} = \omega^2 LC - 1 \tag{2-13a}$$

6. Garstang, W. W., Electronics, Feb. 1932, p. 50; Waidelich, D. L., Electronics, May 1941, p. 28.

The value of the product LC required to give a smoothing factor of value α is

$$LC = \frac{\alpha + 1}{\omega^2} \tag{2-14}$$

Under the same assumption, the value of LC required for each stage of a filter having n similar sections is approximately

$$LC = 0.0253 \frac{\sqrt[n]{\alpha_1} + 1}{f_r^{\,2}} \tag{2-15}$$

in which α_1 is the required smoothing factor at the ripple frequency f_r.

It can be seen from eq. (2-15) that a desired value of smoothing factor may be obtained with a single-section filter or with two or more sections. An infinite number of combinations of L and C may also be used to give the required value of the product LC. The choice of the number of sections and of the values of L and C is influenced by such considerations as available standard parts, relative cost of condensers and chokes, weight and size of condensers and chokes, and increased voltage regulation resulting from choke resistance. All sections of a multisection filter may usually be identical.

Although there is great latitude in the values of L and C that may be used with a given product LC, there is a lower limit to the inductance L_1 of the first section of the filter, below which it is not desirable to go. When the inductance of the first choke is insufficient to maintain a continuous flow of current into the filter, the direct output voltage varies appreciably with load. It is also desirable to maintain continuous flow of current in order to keep the peak anode current as low as possible and to make the tube and transformer efficiency high. Low peak anode current and low voltage regulation are obtained if the inductance L_1 of the first choke exceeds the value

$$L_o = 0.18 \frac{\rho R_t}{f_r} \tag{2-16}$$

at minimum load and twice this value at full load, where R_t is the total resistance of the load, including the chokes. For a 60-cycle, single-phase full-wave rectifier, eq. (2-16) reduces to

$$L_o = R_t/1000 \tag{2-17}$$

At intermediate values of load, the required minimum first choke inductance may be assumed to vary linearly with load current. Since the flow of direct current through the chokes saturates the cores and thus reduces the inductance, the air gap must be of such length as to ensure that the actual value exceeds the required value at all loads. So-called "swinging chokes" are designed to meet this requirement. Too large an air gap is likely to result in insufficient inductance at light loads; too small a gap may result in insufficient inductance at intermediate loads, which may in turn lead to violent "motorboating" (low-frequency oscillation) of the filter.[7]

2-17. Resistance-Condenser Filters.- In certain applications, the chokes of the smoothing filter of Fig. 2-5b may be replaced by resistors. The d-c resistance of resistance-condenser filters is much higher than that of inductance-condenser filters using condensers of equal capacitance and having the same smoothing factor. This higher resistance results in poorer voltage regulation, lower terminal voltage, and greater heat dissipation. The peak anode current is also higher, and the transformer efficiency lower. Resistance-condenser filters incorporating small low-cost electrolytic condensers of high capacitance may sometimes be used to advantage in applications in which low cost, lightness, and small size are primary considerations and in which the current demand is small, as in power supplies used to supply bias voltages or anode voltages for cathode-ray tubes.

2-18. Inductance-Condenser Filters with Condenser Input.- In some applications it is advantageous to use a condenser in the input of an inductance-capacitance filter, as shown in Fig. 2-5c. The condenser greatly reduces the ripple and raises the direct voltage output to a value that approximates crest secondary voltage at light load. As has already been pointed out, however, the increase of the rate at which the condenser discharges between charging periods causes the direct voltage to fall rapidly with increase of load current. Because anode current flows only during short portions of the cycle, as shown in Fig. 2-6, the ratio of the peak anode current to the load current is high, and the efficiency is low. The high peak anode current is particularly objectionable in mercury-vapor rectifier tubes, since it may result in damage to the cathode.

At light loads the ripple in the voltage of the input condenser may be calculated approximately by the use of eq. (2-12). Equation (2-15) may then be used to design the remainder of the filter for a given value of allowable ripple in the output. At high load currents such a filter can be designed accurately only by methods of successive approximation.[8] For conservative design the effect of the input condenser upon the ripple may be neglected.

---

7. Dellenbaugh, F. S. Jr., and Quinby, R. S., QST, Feb. 1932, p. 14; March 1932, p. 27; April 1932, p. 33.
8. Stout, M. B., Electrical Engineering, 54, 977 (1935).

2-19. Choice of Tubes for Power Supplies.- Because mercury-vapor tubes have lower voltage drop, they give higher efficiency of operation than high-vacuum tubes. Because the voltage drop is practically constant in mercury-vapor tubes throughout the normal range of current, they give better voltage regulation than high-vacuum tubes. Mercury-vapor tubes pass higher current than high-vacuum tubes of equal size. The voltage that can be rectified by a mercury-vapor tube is however limited because of danger of glow and subsequent breakdown into an arc during the time the anode is negative. Mercury-vapor tubes are more readily damaged as the result of excessive anode current, and adequate provision must be made to prevent application of anode voltage before the cathode temperature has reached the operating value. The abrupt changes of anode current which take place when mercury-vapor tubes build up or break down produce high-frequency transient disturbances that may be objectionable in some applications of rectifiers. These disturbances must be prevented from affecting other circuits connected to the power supply and from being transferred to the a-c line or radiated.

The complete circuit of a single-phase half-wave power supply is shown in Fig. 2-8. In order to prevent the by-passing of ripple voltage around the chokes as the result of interwinding capacitance to ground, the chokes should always be placed in

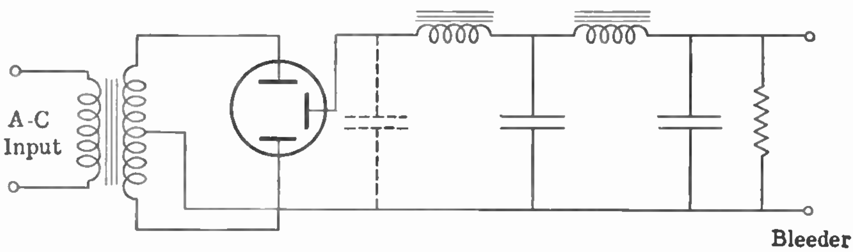


FIG. 2-8.

the d-c line that is not grounded. Since in most applications of power supplies the negative side of the d-c output is grounded, the chokes are usually placed in the positive line. To prevent ripple resulting from the picking up of stray a-c fields by the last choke, care must be exercised in arranging the component parts of the power supply. The final choke should in general not be placed too close to the power transformer, and should be turned so that there is a minimum of flux linkage.

Several difficulties may result from the fact that the filter may comprise one or more potentially resonant circuits.

If series resonance takes place between the inductance and capacitance of a filter section at a frequency contained in the ripple, the ripple output of the section may be much larger than the input. This difficulty can be prevented by making the resonance frequency of each section lower than the fundamental ripple frequency. Transient oscillations of large amplitude may result from sudden changes of load current, or when the power is first applied to the primary of the transformer. One function of the bleeder resistance across the output of the filter is to prevent breakdown of filter condensers as the result of high transient voltages. This resistor may also serve as a voltage divider.

Care must always be taken to ensure that the peak inverse voltage (i.e., greatest magnitude of negative potential of anode with respect to cathode or other electrode) to which the rectifier tubes are subjected does not exceed the allowable value specified by the manufacturer. The peak inverse voltage is greatest when a condenser is used in the input to the filter, being approximately twice the crest secondary voltage in the circuit of each anode. Values for resistance and inductive load are given in Table 2-I. It is extremely important that high-voltage and filament transformers be adequately insulated.

In order to prevent the possibility of oscillation in a multi-stage amplifier with which it is used, the terminal impedance of a power supply should be low at the low-frequency end of the frequency band passed by the amplifier. Since the terminal impedance is determined almost entirely by the capacitance of the final condenser, this condenser must be made large enough so that its reactance is less than the allowable terminal impedance.

2-20. Illustrative Problem.- Application of the principles discussed above can best be illustrated by the design of a single-phase full-wave rectifier that will furnish a direct voltage of 1500 volts at a load current of 250 milliamperes and a ripple factor not to exceed 0.1 per cent.

Examination of Table 2-I shows that the ripple factor at the output of the rectifier is 0.667, that the ripple frequency is 120 cycles, and that the peak inverse tube voltage is 3.14 x 1500 = 4700 volts. Reference to a transmitting-tube manual shows that the type 866-A/866 mercury-vapor rectifier tube can carry the required current and withstand the peak inverse voltage. Substitution of $\alpha = 0.667/0.001 = 667$ and $f_r = 120$ cps in eq. (2-14) gives the following values of LC:

Single-section filter......1174×10^{-6} henry x farad
Two-section filter.........47.1×10^{-6} henry x farad
Three-section filter.......17.1×10^{-6} henry x farad

It is evident from these figures that a two-section filter can give the required filtering without the use of excessively large condensers. Assume that a 50,000-ohm bleeder will be used and that the choke resistance is 75 ohms per choke. The bleeder current is 30 milliamperes, and R_t at full load is $1500/(250 + 30)10^{-3} + 150$, or 5520 ohms. L_1 should therefore exceed 5520/500 or approximately 11 henrys at full load. Under the assumption that choke inductance will be reduced approximately 50 per cent from no load to full load, a first choke having a no-load value of 25 henrys will be satisfactory at full load. Since at no load R_t is equal to the resistance of the bleeder plus that of the chokes, eq. (2-17) calls for a no-load inductance of approximately 50 henrys for the first choke. This value can be reduced to approximately 25 henrys by the use of a 25-ohm bleeder. It is unlikely, however, that the importance of good voltage regulation at very light loads is great enough to offset the disadvantages of the higher current drain of the lower-resistance bleeder and the accompanying higher heat dissipation. For this reason the 50,000-ohm bleeder and 25-henry choke should be satisfactory. The required capacitance is then 3.77 microfarads. A two-section filter using 25-henry chokes and 4-microfarad condensers will give the required filtering. The resonant frequency of each section is $1/2\pi\sqrt{12.5 \times 4 \times 10^{-6}} = 22.5$ cps at full load, which is well below the ripple frequency. At full load the voltage drop through the chokes is $150 \times 0.280 = 42$ volts. The tube voltage drop is 15 volts. Table 2-I shows that the ratio of the rms secondary voltage per tube to the direct voltage output of the rectifier, neglecting tube drop, is 1.11. Therefore the rms secondary voltage per tube should be $1.11\ (1500 + 42 + 15) = 1730$ volts at full load. The terminal impedance is approximately equal to the reactance of the second filter condenser, which is roughly 400 ohms at 100 cycles.

2-21. Voltage Stabilizers.- In many of the circuits discussed in later chapters of this book it is necessary to use high voltages that are independent of fluctuations of supply voltage and of load. Direct voltages of almost any desired degree of constancy can be obtained by the use of electronic voltage stabilizers. Since a stabilizer that eliminates random changes of voltage will in general also eliminate periodic changes of voltage, voltage stabilizers used in the output of a power supply produce the additional beneficial effect of reducing ripple voltage. Types of stabilizers that reduce the variation of voltage with load current also reduce the internal terminal impedance of the stabilized voltage supply.

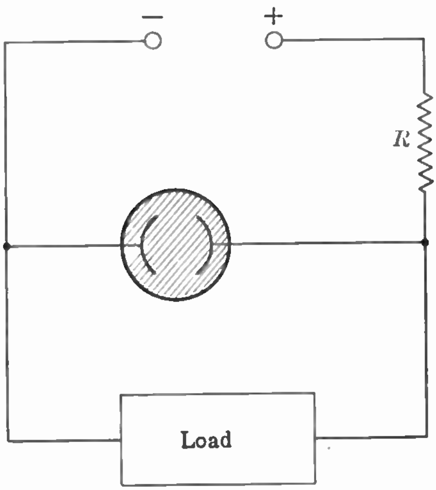


FIG. 2-9. Glow-tube voltage stabilizer.

The simplest type of voltage stabilizer consists of a glow-discharge tube used with a resistance, as shown in Fig. 2-9. Over a wide range of current, called the "normal" current range, the voltage of a glow-discharge tube having a cathode of large area is nearly independent of current. Increase of input voltage in the circuit of Fig. 2-9 results in an increase of IR drop through the series resistance, the tube voltage increasing only slightly. Increase in current through the load results in an almost equal reduction in tube current at nearly constant tube voltage. Tubes such as the 874, the VR-75-30, the VR-105-30, and the VR-150-30 are designed to operate at different useful values of voltage. Where higher voltages are required, two or more tubes may be used in series, in which case several regulated voltages may be obtained from the same circuit, as shown in Fig. 2-10. The series resistance and input voltage must always be such that the tubes fire and the tube currents lie within the range for which the tubes are designed, throughout the entire range of load current desired.

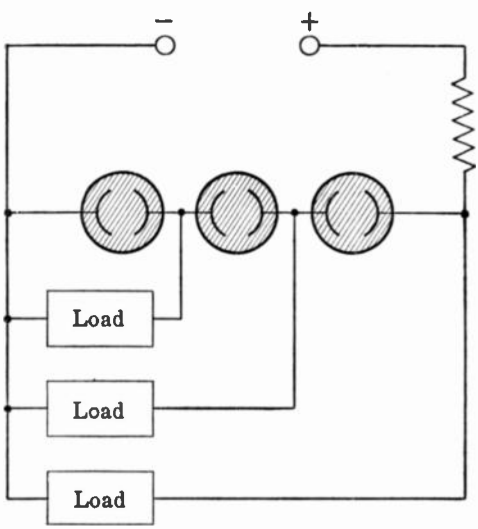


FIG. 2-10. Glow-tube voltage stabilizer.

Figure 2-11 shows a high-vacuum-tube circuit that is essentially a degenerative amplifier (see Sec. 3-11). The resistance R may be the resistance of the load. Increase of output voltage E_0 as the result of either increase of load resistance or increase of input voltage increases the negative grid bias and thus tends to reduce the

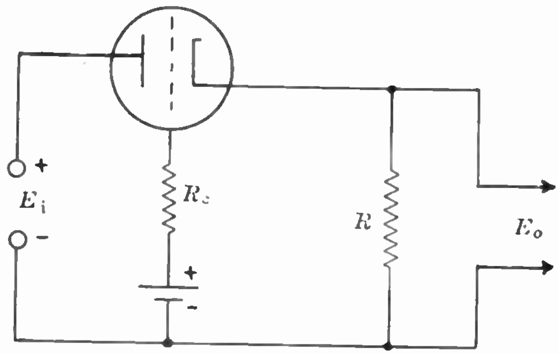


FIG. 2-11.

plate current. The reduction of plate current in turn partially eliminates the change in output voltage. The purpose of R_c is to prevent the flow of excessive grid current when the input voltage is removed. The analysis of the equivalent plate circuit given in Sec. 3-11 shows that high stabilization necessitates the use of tubes with high amplification factor. To prevent loss of output voltage resulting from voltage drop through the tube, high-current tubes should be used. Since tubes having high amplification factor in general pass low plate current, this circuit is useful only in applications where the load current is small.

The circuit of Fig. 2-12a is based upon the bridge circuit used for the measurement of amplification factor.[9] If R_2 is equal to μR_1, the increase of negative grid voltage resulting from increase of E_i is compensated by the increase of plate voltage and the plate current does not change. Hence the voltage across R_3, which may be the load, remains constant throughout any range of voltages in which the amplification factor is essentially constant. By making R_2 slightly smaller than μR_1, the output voltage can be made to increase with decrease of input voltage. The output voltage is less than the input voltage by the sum of the voltage across R_1 and the drop through the tube. Since R_1/R_2 decreases with increase of μ, the small voltage loss in R_1 requires the use of a tube with high amplification factor. This circuit, like the degenerative circuit, is therefore best suited to applications in which the load current is small.

The circuit of Fig. 2-12b is based upon a bridge circuit used in the measurement of transconductance.[10] Increase of E_i

---

9. See Standards on Electronics, 1938, Inst. of Radio Eng., p. 32.
10. See Standards on Electronics, 1938, Inst. of Radio Eng., p. 30.

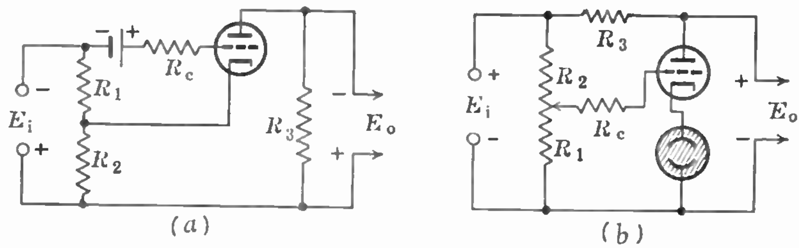


FIG. 2-12. Mu-bridge and transconductance-bridge voltage stabilizers.

makes the grid less negative and thus increases the plate current. By proper choice of circuit constants, the resulting increase of IR drop across R_3 can be made to equal the increase of input voltage, so that the output voltage remains constant. Analysis of the equivalent plate-circuit shows that this result is attained when $R_3 = (R_1 + R_2)/R_1 g_m$. Since R_3 should be small in order to prevent unnecessary loss of voltage, a high-transconductance tube should be used in this circuit. As in the circuits of Figs. 2-11 and 2-12a, the resistance R_c prevents the flow of high grid current in case the circuit is improperly adjusted. The glow tube furnishes a constant biasing voltage.

Neither of the circuits of Fig. 2-12 compensates for changes of output voltage resulting from changes of load. In fact, the series resistances make the voltage regulation poorer and increase the effective internal impedance of the voltage supply. Figure 2-13 shows a circuit that compensates for both changes of input voltage and of load current and has a low internal impedance between the output terminals. This circuit, which is the most commonly used type, is essentially a two-stage inverse-feedback circuit. Increase of E_o decreases the size of the negative grid voltage of T_1 and thus increases the plate current of T_1. The resulting increase of voltage across R_2 makes the grid of T_2 more negative thus reducing the plate current of T_2 and hence tending to reduce the output voltage. The output voltage may be varied by means of R_1. Because of the voltage-dividing action of this resistance, only a portion of the output voltage change is applied to the grid of T_1. Addition of the condenser C increases the fraction of the voltage change impressed upon the grid of T_1 when the change occurs rapidly. This condenser thus increases the filtering action of the circuit, but may lead to undesirable transient oscillations of the output voltage because the initial compensation as the result of a change of output voltage exceeds the final value obtained after the condenser voltage reaches equilibrium. For this reason the capacitance should not be too high. The coupling resistor R_2 should be large in comparison with the plate resistance of T_1. R_3 should be chosen so as to limit the

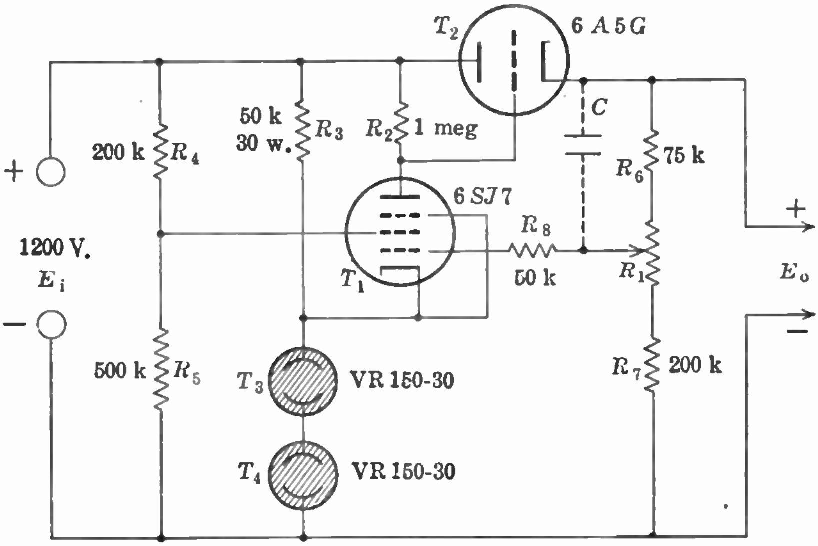


FIG. 2-13. Two-stage voltage stabilizer.

current through the glow tube to the range for which it is designed. Although the screen voltage for T_1 may be taken from the output side of the circuit, the variation of screen voltage helps to stabilize the output voltage when the screen voltage is taken from the input side, as shown in Fig. 2-13. Because of the large difference of potential between the cathodes of T_1 and T_2, it is necessary to use separate heater transformers, insulated against the high voltage.

2-22. Current Stabilizer.- In certain circuits, notably in magnetron oscillators, (see Sec. 10-14), it is essential to

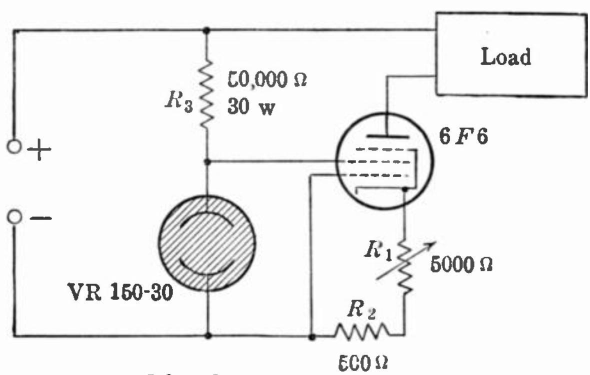


FIG. 2-14. Current regulator.

maintain a constant current. Figure 2-14 shows a circuit that accomplishes this result. The circuit makes use of the fact that in certain ranges of voltages and plate current the plate current of pentodes varies comparatively little with plate voltage. Additional stabilization results from the use of a cathode biasing resistance, which increases the grid bias with increase of plate current. The circuit constants given in Fig. 2-14 are suitable for an input voltage ranging from 200 volts to 1200 volts. The cathode resistor R_1 varies the current from about 5 to 30 milliamperes. R_3 must be large enough to prevent the flow of excessive screen current when the plate current is reduced to a low value by opening the plate circuit or increasing the load resistance to a high value. The glow tube may extinguish when the plate current is reduced by the load to a value below that normally maintained by the circuit, but this is not objectionable. In the vicinity of 20 milliamperes, the plate current varies by about $1\frac{1}{2}$ ma when the input voltage is varied from 300 to 1000 volts or when the output is short-circuited.

Chapter 3

AMPLIFICATION

An electrical amplifier is a device used for effectively increasing the voltage or power of a source of energy, through the control by the input power of a larger amount of power supplied by a local source to a load or energy-consuming element. All electrical amplifiers require a d-c source of energy from which the energy delivered to the load is derived. When a-c power is delivered to the load, the amplifier functions as a power converter and in most instances the power from the original a-c source is used only to control the amount of power converted, and only a small amount of energy flows from the original source to the load. This is in contrast with an ordinary electrical device such as a transformer which may also be used for a voltage amplifier or to increase the power in a load by making the source deliver more power than it would without the transformer.

The high-vacuum thermionic tube is for many purposes superior to all other known electrical devices for performing the task of voltage or power amplification, because of the manner in which the grid of the tube controls electrons in their journeys from the cathode to the plate without capturing any of the electrons. When the grid is always negative with respect to the cathode, the vacuum-tube amplifier can be used with almost any kind of a source, i.e. with sources having a very wide range of impedances and power outputs. A vacuum-tube amplifier may consist of as many tubes arranged in cascade as it is practical to use, being limited by electrostatic and electromagnetic coupling between the final output and the input, and by the fact that the flow of electrons in both evacuated space and solid conductors fluctuates somewhat and thereby generates minute pulsating voltages[1] which may be of the same order of magnitude as the voltage to be amplified. There are also other things, such as mechanical vibration, that limit the amount of amplification obtainable by the use of vacuum tubes. These limits on amplification are being reduced as new methods for overcoming some of the troubles are gradually developed.

3-1. Amplifier Classifications.- Vacuum-tube amplifiers may be classified in the following several different ways: mode

1. In some cases of the order of a microvolt.

of operation, commercial application, and output power.

One classification relates the mode of operation of the tube with respect to its d-c and a-c grid and plate voltages and currents. Under this classification are Class A, Class AB, Class B and Class C amplifiers.

A Class A amplifier operates with such d-c grid and plate potentials and such a-c input, or grid, voltage that the a-c plate current flows for the full cycle of the input voltage and has substantially the same wave shape as that of the input voltage for any kind of an output impedance. Ordinarily the grid of a Class A amplifier is not driven positive with respect to the cathode. Figure 3-1 illustrates the operation of a Class A amplifier with respect to the I_b - E_c characteristics of the tube and the nature of the resulting a-c plate current.

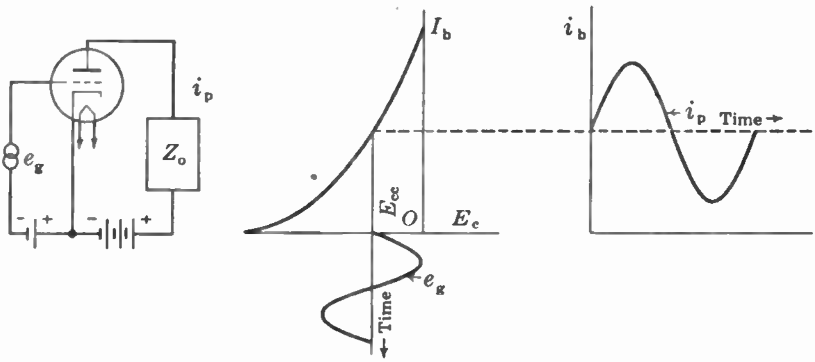


FIG. 3-1. Mode of operation of a Class A amplifier. Z_o is the output impedance or load impedance.

A Class AB amplifier operates with such d-c grid and plate potentials and such a-c input voltage that the a-c plate current flows for appreciably more than half but less than the entire cycle of the input voltage. The grid of a Class AB amplifier may be driven positive with respect to the cathode. Class AB_1 means that the grid is not driven positive whereas in the AB_2 type the grid is driven positive. Figure 3-2 illustrates the operation of the Class AB amplifier with respect to the I_b - E_c characteristics of the tube and the nature of the plate current of a single tube. The plate current is badly distorted but by the use of two tubes in a push-pull circuit (which will be treated later), the output current of the system will have nearly the same wave shape as the input voltage.

A Class B amplifier operates with d-c grid potential substantially at plate current cutoff for the operating value of d-c plate potential. Approximately, plate current flows only

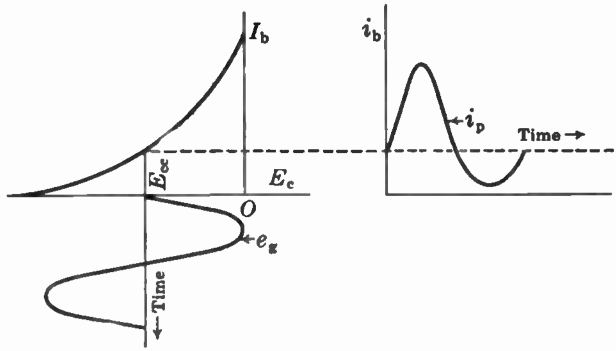


FIG. 3-2. Mode of operation of a Class AB amplifier tube.

during the positive half of the input cycles. When the output impedance has no marked selective properties it requires two tubes in a push-pull arrangement with a simple transformer to produce an output current that is substantially like the input voltage. When the output impedance has selective properties (for example, when it is a parallel resonant circuit) the output current may have nearly the same wave shape as the input voltage with but a single tube. Figure 3-3 illustrates the operation of the Class B amplifier with respect to its I_b - E_c

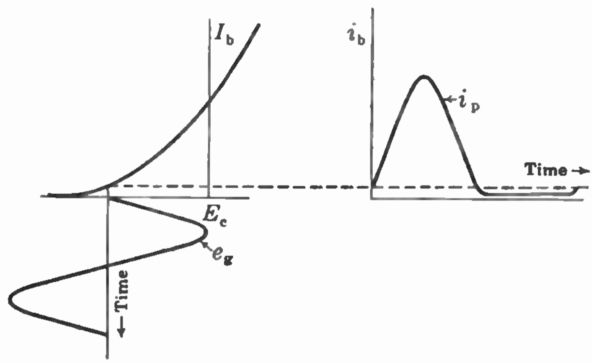


FIG. 3-3. Mode of operation of a Class B amplifier.

characteristics, and the nature of the plate current. The grid of a Class B amplifier is usually driven positive with respect to the cathode.

A Class C amplifier operates with d-c grid potential more negative than plate current cutoff for the d-c operating plate potential. Plate current flows for less than one-half of the a-c input cycle. Because of the nature of the plate current it is necessary to use a selective circuit for the output impedance of this type of amplifier to get nearly distortionless power output. This type of amplifier is used only when the a-c input voltage is constant. Figure 3-4 illustrates the mode of operation of the Class C amplifier with respect to the I_b - E_c characteristics of the tube. The grid of a Class C amplifier is

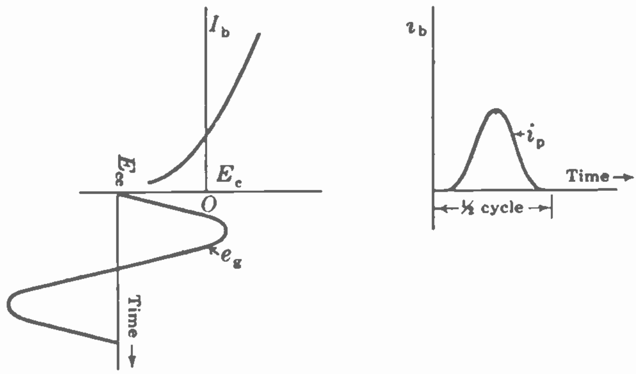


FIG. 3-4. Mode of operation of a Class C amplifier.

generally driven positive with respect to the cathode.

The second classification of amplifiers refers to their application as audio-frequency, video-frequency, radio-frequency or direct-current amplifiers. An audio-frequency amplifier is a wide-band or non-selective amplifier used for amplifying in the audio or low region of the frequency spectrum. In general the amplification does not change rapidly for frequencies below and above the range of interest. A video-frequency amplifier is a very wide band amplifier used, generally, for amplifying television signals. For this purpose the gain of the amplifier must be substantially constant from very low frequencies up to a value of a few megacycles or more. A radio-frequency amplifier is usually a narrow-band tuned, or selective, amplifier used in the radio-frequency spectrum. The frequency range over which the amplification or power output is substantially constant for a constant-voltage input is only a small percentage of the mid-operating frequency. A direct-current amplifier is a type of amplifier used for amplifying direct currents or voltages, usually containing frequencies not in excess of a few cycles per second, i.e. so low that it is necessary to design the amplifier to respond to zero frequency also.

A third classification of amplifiers refers to their output power. When a tube is used for the purpose of converting d-c power from the d-c source in the plate circuit into a-c power at a reasonable efficiency, it is classed as a power amplifier.

If on the other hand, a tube is used to increase the voltage from the input to the output, and the power converted and efficiency are of no consideration, it is classed as a voltage amplifier. There is no sharp distinction between a power amplifier and a voltage amplifier. A multi-stage amplifier may contain both voltage and power amplifier stages. Usually the input stages of a general amplifier increase the voltage from a source to an amount necessary to drive the tube furnishing power to the utilization device. There are also certain applications of amplifiers where current amplification is desirable. An example of this is an amplifier employed between a photo tube and a relay. The individual stages of the amplifier may be voltage amplifiers but the final tube is used to produce current in the relay.

PART I. AUDIO-FREQUENCY VOLTAGE AMPLIFIERS

3-2. Single Tube Class A Amplifier.- The circuit diagram for a single tube Class A amplifier is shown in Fig. 3-5. The expression for the voltage amplification for this circuit is

$$\underline{A} = \frac{\underline{E}_o}{\underline{E}_i} = \frac{-\mu \underline{Z}_o}{r_p + \underline{Z}_o} \tag{3-1}$$

where $\underline{E}_i$ is the symbol for the complex alternating input voltage (cathode to grid), and $\underline{E}_o$ that of the output voltage. This expression also applies to a pentode tube when the screen and suppressor are operated at zero a-c potential with respect to the cathode.

When $\underline{Z}_o$ is a pure resistance R_o the expression becomes

$$\underline{A} = \frac{-\mu R_o}{r_p + R_o} = \frac{-\mu}{\frac{r_p}{R_o} + 1} \tag{3-2}$$

The negative sign indicates a 180° phase displacement between

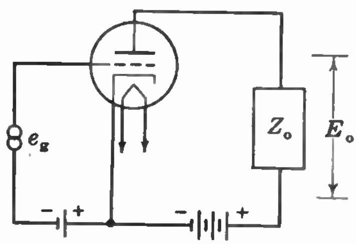


FIG. 3-5. Simple Class A amplifier.

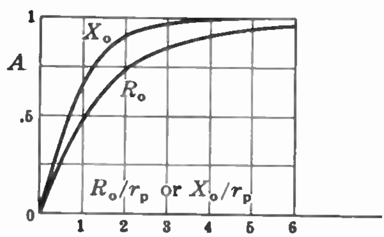


FIG. 3-6. Voltage amplification Z_o/r_p for a Class A amplifier.

the output voltage $\underline{E}_o$ and the input voltage $\underline{E}_i$. The relation between A and R_o/r_p is shown by the curve labelled R_o in Fig. 3-6. This curve shows that R_o/r_p must be several times greater than unity in order that the voltage amplification approach the amplification factor of the tube. Such a relationship is generally achieved in ordinary practice only when the tube has a plate resistance less than 50,000 ohms because of the high d-c voltage required in the plate source to make up for the potential drop across R_o. Small pentode amplifier tubes are usually operated with a resistance R_o that is only a fraction of r_p. This makes R_o/r_p less than unity and in this range the voltage amplification is almost directly proportional to R_o. Hence for a pentode Class A amplifier when $r_p/R_o >> 1$

$$\underline{A} = \frac{-\mu}{r_p/R_o} = -g_{gp}R_o \tag{3-3}$$

When the output impedance $\underline{Z}_o$ is a highly inductive reactance such that $X_o >> R_o$ the expression for $\underline{A}$ becomes

$$\underline{A} = \frac{-\mu j\omega L_o}{r_p + R_o + j\omega L_o} \tag{3-4}$$

and

$$A = \frac{\mu\omega L_o}{\sqrt{(r_p + R_o)^2 + \omega^2 L_o^2}} = \frac{\mu}{\sqrt{\left[\dfrac{r_p + R_o}{\omega L_o}\right]^2 + 1}} \tag{3-5}$$

For this case the relation between the voltage amplification and the ratio of ωL_o to $r_p + R_o$ is given by the curve labelled X_o of Fig. 3-6. For the pure resistance case the voltage amplification is independent of frequency for all frequencies for which the output impedance is substantially a pure resistance. For the highly inductive reactance case the voltage amplification depends upon frequency, but becomes substantially independent of frequency when

$$\frac{r_p + R_o}{\omega L_o} << 1$$

There is a phase shift θ from 180° which also depends upon frequency. The phase shift θ is given by the approximate expression

$$\theta = \tan^{-1}\frac{r_p}{\omega L_o} \tag{3-5a}$$

and leads the 180° displacement. For many applications of the amplifier this change of phase shift with frequency is more serious than the change in voltage amplification.

When the output impedance is made up of a resistance R_o and capacitance C_o in parallel the voltage amplification becomes

$$\underline{A} = \frac{-g_{gp}}{g_p + G_o + j\omega C_o} \tag{3-6}$$

$$A = \frac{g_{gp}}{\sqrt{(g_p + G_o)^2 + \omega^2 C_o^2}} = \frac{g_{gp}}{g_p + G_o}\frac{1}{\sqrt{1 + \left(\frac{\omega C_o}{g_p + G_o}\right)^2}}$$

where[2] $g_{gp} = \frac{\mu}{r_p}$, $g_p = \frac{1}{r_p}$ and $G_o = \frac{1}{R_o}$ The phase shift θ is

$$\theta = \tan^{-1}\frac{\omega C_o}{G_p + G_o} \tag{3-7}$$

For this case θ lags the 180° phase displacement between $\underline{E}_o$ and $\underline{E}_1$. When either the frequency or capacitance is increased the phase shift is increased and the voltage amplification decreased.

More complex impedances than these simple cases will be treated as they arise in connection with the various types of amplifiers. However it is well for the student to have these simple cases in mind because they appear later when treating multi-stage amplifiers.

Effects of Interelectrode Capacitances on Voltage Amplification. The interelectrode capacitances of a triode are indicated in Fig. 3-7. For a pentode Class A amplifier the plate to cathode capacitance includes the direct plate to cathode capacitance plus the plate to suppressor capacitance plus the plate to screen capacitance. These three capacitances are generally lumped together and are called output capacitance. In a similar manner the capacitance between grid and cathode comprises the direct grid to cathode capacitance plus the grid to screen capacitance.

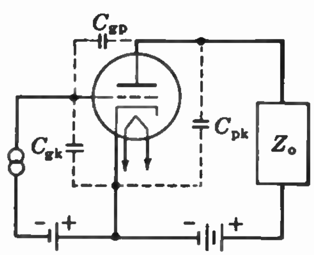


FIG. 3-7. Interelectrode capacitances of a triode (C_{gk}, C_{pk}, C_{gp}).

For either type of tube and a general impedance $\underline{Z}_o = 1/\underline{Y}_o$ in the output circuit the voltage amplification is given by the expression

$$\underline{A} = \frac{j\omega C_{gp} - g_{gp}}{G_o + g_p + j\omega C_{gp} + j\omega C_{pk} + jB_o} \tag{3-8}$$

where $G_o + jB_o$ is the output or load admittance $\underline{Y}_o$, equal to $1/\underline{Z}_o$. For most all applications of Class A amplifiers the quantity $j\omega C_{gp}$ in the numerator of expression 3-8 can be neglected.

---

2. According to I.R.E. standards either of the symbols g_m and g_{gp} may be used to designate μ/r_p (mutual conductance or grid-plate transconductance); g_m was used in Ch. 2.

Hence the main effect of the interelectrode capacitances on voltage amplification is due to the manner in which these capacitances modify the output impedance of the amplifier.

3-3. **The Input Admittance of a Class A Amplifier.**- When the grid of a high vacuum tube is operated with sufficient negative bias to prevent the grid from becoming positive with respect to the cathode under dynamic operation the grid conductance[3] g_g is very small and may be ignored for most practical applications. However, because of the interelectrode capacitances, first the grid circuit requires some charging current, and second, the grid-to-plate capacitance provides coupling between the plate circuit, with its higher energy level, and the input circuit. The latter effect may become most troublesome as will be seen by the following analysis.

Referring to Fig. 3-7 it is seen that the input admittance of a triode consists of two parallel paths. Path one is the capacitance C_{gk} between grid and cathode. Path two is the capacitance between grid and plate C_{gp} in series with the output admittance. Through path two the input voltage and output voltage operate in series vectorially to produce a current. The current $\underline{I}_2$ through path two is $\underline{I}_2 = (\underline{E}_g - \underline{E}_p)\, j\omega C_{gp}$. But $\underline{E}_p = \underline{A}\,\underline{E}_g$. Hence $\underline{I}_2 = (\underline{E}_g - \underline{A}\,\underline{E}_g)\, j\omega C_{gp} = \underline{E}_g\,(1 - \underline{A})\, j\omega C_{gp}$. Consequently the total input current is $\underline{I}_1 = \underline{E}_g\,[j\omega C_{gk} + j\omega C_{gp}\,(1 - \underline{A})]$ and the input admittance is $\underline{Y}_1 = \frac{\underline{I}_1}{\underline{E}_g} = j\omega C_{gk} + j\omega C_{gp}\,(1 - \underline{A})$. The voltage amplification $\underline{A}$ can be expressed in the form $\underline{A} = A \angle\varphi$ or $\underline{A} = A\,[\cos\varphi + j\sin\varphi]$. Substituting this form for $\underline{A}$ in the above expression for $\underline{Y}_1$ there results

$$\underline{Y}_1 = j\omega C_{gk} + j\omega C_{gp}\,(1 - A\cos\varphi - jA\sin\varphi) \qquad (3\text{-}9)$$
$$= \omega C_{gp}\,A\sin\varphi + j[\omega C_{gk} + \omega C_{gp}\,(1 - A\cos\varphi)]$$

where φ is the phase displacement between the a-c grid and plate voltage.

When the output impedance of the amplifier is a pure resistance of a value that is considerably smaller than that of the interelectrode (capacitive) reactance, φ is 180° and the input admittance becomes a capacitive susceptance equal in value to

$$\omega C_{gk} + \omega C_{gp}\,(1 + A).$$

3. Grid conductance is the ratio of the in-phase component of the grid alternating current (sine-wave) to the grid alternating voltage, all other electrode voltages being maintained constant. Effective grid conductance may become important at high frequencies, see Sec. 10-1.

For triode amplifiers where A is generally several times greater than unity and C_{gp} and C_{gk} are about equal to each other, the input admittance is due largely to the grid to plate capacitance C_{gp}. For pentode amplifiers where C_{gp} is less than 0.5 per cent of C_{gk} and practical voltage amplifications are rarely greater than 100, the input admittance is due largely to the grid to cathode capacitance C_{gk}. Hence the input admittance of a pentode amplifier is much smaller than that of a corresponding triode. Consequently when the source has a comparatively high impedance a pentode amplifier should be used because the voltage delivered to the grid of the tube will be higher and will change less with frequency.

The phase displacement angle φ of an amplifier is greater than 180° and less than 270° when the output impedance is inductive. Sin φ is then negative. The term ωC_{gp} A sin φ in eq. (3-9) is negative. Hence with an inductive output impedance, the input admittance is a negative conductance plus a positive susceptance. This fact means that some of the output power is fed back into the input circuit in such a phase relation that there is a component of input current 180° out of phase with the input voltage. This occurrence is often troublesome when a triode is used at high frequencies with a tuned circuit in the output, because for frequencies for which the output circuit is inductive the input conductance is negative. A negative input conductance may cause oscillation or regeneration and thus render the amplifier useless. For pentode tubes the quantity ωC_{gp} A sin φ is usually so small that, if proper care is taken to shield the input circuit from the output, appreciable regeneration or oscillation will occur only at quite high frequencies. The negative input conductance of an amplifier can be overcome by the expedient which is known as the neutralization of the effect of the grid to plate capacitance. Neutralization is accomplished in one method by obtaining a voltage in the output circuit that is 180° out of phase with the plate to cathode potential and impressing this voltage on the grid of the tube through a condenser about equal in value to C_{gp}; thereby a current is impressed in the input circuit in phase with the input voltage. When the current so produced is just equal to that caused by the grid-to-plate capacitance C_{gp} of the tube the neutralization is complete. More details about neutralization will be given in the section on tuned power amplifiers. (Sec. 3-22)

3-4. Distortion in Amplifiers.- For most of the applications of a vacuum tube amplifier the input voltage to the grid of the tube is not a simple sine function, but is generally made up of several harmonic functions. For a complex voltage of this kind to be amplified without distortion the amplifier must be free from amplitude, frequency and phase distortion. All three of these distortions are present to some extent in vacuum tube amplifiers but can be reduced to a satisfactory degree by careful design and certain expedients which will be discussed in later sections.

Amplitude distortion exists when the output voltage of an amplifier contains frequencies not present in the input voltage over the complete range of the input voltage. Non-linearity between output and input voltages is caused by the fact that the tube factors μ and r_p, particularly r_p, are not constant over the range of operation. These two factors were treated as constants in the discussion of Sec. 3-2 for the purpose of arriving at the approximate relation between output and input voltages. When amplitude distortion is present the output voltage contains harmonic voltages not present in the input voltage. Hence the output wave form is not an exact replica of the input wave form. This kind of distortion depends largely upon the nature of the various volt-ampere characteristics of the tube in the region over which the input and output voltages range and can be reduced by limiting the range of the two voltages to a region in which there is nearly linearity.

Frequency distortion exists when all frequencies of an input voltage are not amplified at the same rate. An amplifier with reactance in the output circuit gives very bad frequency distortion when the output impedance is not large compared to the plate resistance. This is illustrated by expression (3-5). An amplifier with a tuned circuit in the output gives the worst kind of frequency distortion. Such amplifiers are used purposely when it is desired to amplify only a very small band of frequencies.

Phase distortion in an amplifier exists when the relative phases of the various frequency components in a complex input voltage are not preserved in the output voltage. Phase and frequency distortion in the plate voltage of an amplifier generally exist together, i.e. when an amplifier has frequency distortion it also has phase distortion and vice versa. Phase distortion usually causes no reduction in quality when the amplifier is used in sound work[4] but it is sometimes troublesome in this field when an attempt is made to reduce frequency distortion by means of negative feed-back. For television amplifiers, commonly known as video amplifiers, phase distortion is a troublesome thing and much emphasis is placed on methods that will reduce it. Phase distortion is also undesirable in amplifiers used with cathode ray oscilloscopes.

3-5. Classification of Audio-Frequency Voltage Amplifiers.- An audio-frequency voltage amplifier is an amplifier,

4. Because the ear does not distinguish moderate changes in relative phase.

generally multi-stage, that is used for amplifying in the frequency spectrum from a few cycles per second up to about 20,000 cycles per second. There are four main types of audio-frequency voltage amplifiers. These types refer to the manner in which tubes in a cascade arrangement are coupled, each to its succeeding tube. These types are:

1. Impedance-capacitance Coupled (illustrated by Fig. 3-8). The purpose of the coupling capacitance C_c is to provide an a-c

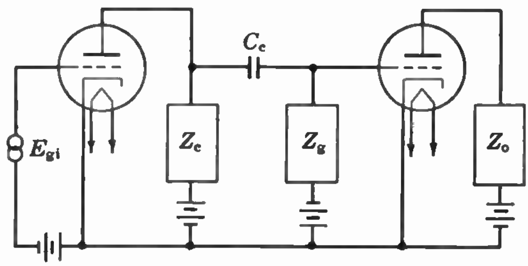


FIG. 3-8. Impedance-capacitance-coupled amplifier.

circuit from the plate of the tube to the grid of the next tube but to keep the d-c plate potential off that grid. The nature of the two impedances Z_c and Z_g is usually limited in practice to two types, namely, pure resistances and highly inductive reactances. This gives two sub-types of coupling, namely,

a. resistance-capacitance coupling, illustrated by Fig. 3-10;

b. inductance-capacitance coupling illustrated by Fig. 3-14.

2. Transformer-coupled Amplifiers. This type is illustrated by Fig. 3-9. The a-c plate voltage of the input tube is

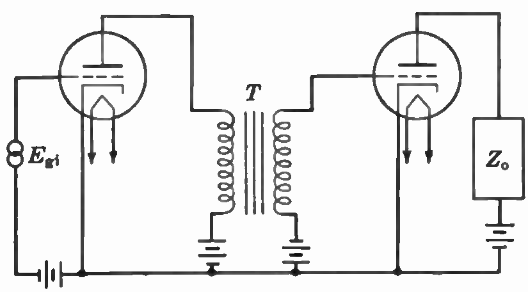


FIG. 3-9. Transformer-coupled amplifier.

transformed to the grid of the output tube and the d-c circuit is broken by using two windings insulated from each other. Some step-up in a-c voltage can usually be obtained by transformer action.

3. Direct Coupled. There are several types of direct-coupled amplifiers, some of which will be discussed later. In all direct-coupled types the main purposes are either to amplify changes in d-c potential or to amplify very low frequencies without amplitude and phase distortion. The abbreviation "d-c amplifier" should be read "direct-coupled amplifier" rather than direct-current, since many d-c amplifiers transmit low but not zero frequency.

The general requirements of an audio-frequency voltage amplifier are:

a. The voltage amplification must remain within certain specified limits between the lowest and highest frequency.

b. The amount of voltage amplification must be sufficient to "raise the source voltage" to the value necessary to drive the load or the power amplifier, respectively.

c. The distortion introduced by the tubes must be within a specified amount usually expressed in per cent of the maximum output voltage.

d. The phase shift must stay within certain limits.

e. The change in gain with changes in d-c operating voltages and changes in filament voltage must be limited.

f. The noise and hum voltages in the output introduced by the tubes and d-c power supply must be limited to a certain percentage of the maximum signal voltage.

g. Mechanical vibration of the tubes must not produce an output disturbance voltage that is greater than a certain per cent of the signal voltage.

3-6. The Resistance-Capacitance-Coupled Amplifier.- The resistance-capacitance-coupled amplifier is illustrated in Fig. 3-10. The resistance R_c is the main coupling resistor and

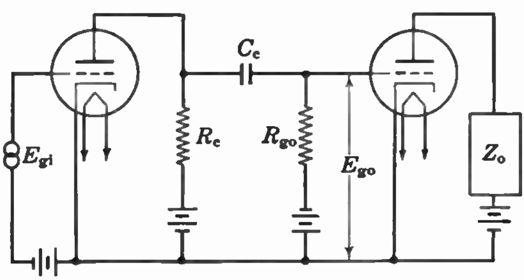


FIG. 3-10. Resistance-capacitance-coupled amplifier.

the voltage developed across this resistor is the plate voltage of the tube. This resistor affects the d-c voltage and the d-c supply voltage must be raised to make up for the drop in R_c. The coupling condenser C_c keeps the d-c plate potential of the input tube off the grid of the output tube. The resistor R_{go} is necessary to bring the grid to a constant d-c potential. R_{go} is usually of the order of 0.5 to 2 megohms for small tubes and should generally be as high as is practical. The properties of this type of coupling are constant amplification over a large range of medium frequencies, but at the low frequencies the amplification falls off because of a drop in potential in the coupling condenser C_c and at the high frequencies it falls because of the interelectrode capacitance of the tube which imposes a small capacitance in shunt with the resistors R_c and R_{go}. Figure 3-11 illustrates a typical curve of amplification versus frequency for this type of amplifier. By the (voltage) amplification is meant the ratio of the grid voltage E_{go} of the output tube to the grid voltage E_{g1} of the input tube, i.e. E_{go}/E_{g1}. This ratio will also be referred to as the gain of the stage. The gain of the stage in decibels[5] is $20 \log_{10} E_{go}/E_{g1}$.

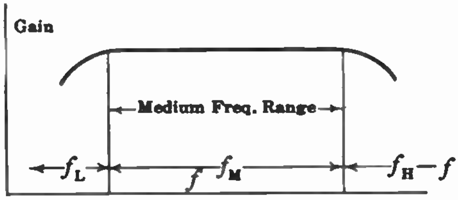


FIG. 3-11. Typical gain vs frequency characteristics of a resistance-capacitance-coupled amplifier.

For purposes of simplification it is convenient to analyze an audio-frequency voltage amplifier by dividing the analysis into three parts, namely, medium frequency, low frequency and high frequency. This is justified because amplifiers for the complete audio-frequency spectrum are so constituted that the gain is substantially constant through the medium frequency range and some departure from an ideal characteristic occurs at the high and low frequencies when there is a compromise between the amount of gain and the manner in which the gain varies with the frequency.

The Medium-frequency Gain. In the medium-frequency range, the gain of a stage of a resistance-capacitance coupled amplifier is determined almost entirely by the constants of the input tube and the sizes of resistors R_c and R_{go} (Fig. 3-10). The voltage drop through the coupling condenser C_c is negligible

5. Rigorously, a decibel is a unit for the measure of power ratio, and only when $E_{go}^2/E_{g1}^2 = P_o/P_1$ is the expression correctly expressed in db, but common practice disregards this distinction.

and the shunting effect of the tube capacitances is negligible. Hence in the medium frequency range the load impedance of the input tube becomes the two resistors R_c and R_{go} in parallel and the a-c voltage on the grid of the output tube is the same as the plate voltage of the input tube. The voltage gain of the stage is (subscript i refers to input tube)

$$\underline{A}_M = \frac{\underline{E}_{go}}{\underline{E}_{gi}} = \frac{-\mu_i R_e}{r_{pi} + R_e} \text{ where } R_e = \frac{R_c R_{go}}{R_c + R_{go}} \tag{3-10}$$

This expression may also be written in the form

$$\underline{A}_M = \frac{\underline{E}_{go}}{\underline{E}_{gi}} = \frac{-g_{gpi}}{g_{pi} + G_c + G_{go}} \tag{3-11}$$

where $g_{pi} = \frac{1}{r_{pi}}$, $G_c = \frac{1}{R_c}$, $G_{go} = 1/R_{go}$ and g_{gpi} is the grid-plate transconductance of the input tube. The phase displacement is 180°.

The Low-frequency Gain.[6] As the frequency is decreased the gain starts dropping from the value given by expression (3-11) when the reactance of the coupling condenser becomes equal to about 0.4 of the resistance R_{go}. Also the load impedance of the input tube starts rising. Hence these two effects change the gain of the stage. The first effect decreases the gain and the second effect slightly raises the gain. The first effect is the larger. The net effect is given by the expression

$$\underline{A}_L = \frac{\underline{E}_{go}}{\underline{E}_{gi}} = \frac{\underline{A}_M}{1 - j\,\frac{G_e}{\omega C_c}} \text{ where } G_e = \frac{G_{go}\,(G_c + g_{pi})}{G_{go} + G_c + g_{pi}} \tag{3-12}$$

The absolute value of $\underline{A}_L$ is

$$A_L = \frac{A_M}{\sqrt{1 + \left(\frac{G_e}{\omega C_c}\right)^2}} \tag{3-13}$$

Expression (3-12) indicates a phase shift from the phase displacement of 180° at medium frequencies. This phase shift θ_L is

$$\theta_L = \tan^{-1} \frac{G_e}{\omega C_c} \tag{3-14}$$

θ_L leads the 180° phase displacement. The reduction in gain in decibels from the medium frequency gain is

$$Db_L = 20 \log_{10} \frac{A_M}{A_L} \tag{3-15}$$

Figure 3-12 shows how the reduction in gain in decibels and the

6. See Sec. 1-14.

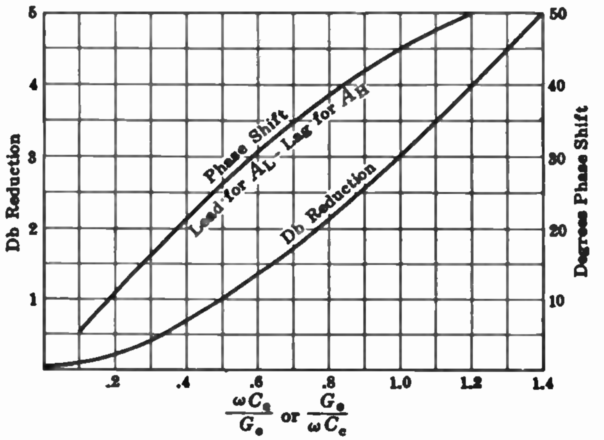


FIG. 3-12. Reduction in gain and phase shift vs $\frac{G_e}{\omega C_e}$ or $\frac{\omega C_e}{G_e}$ in a resistance-capacitance-coupled amplifier.

phase shift depend upon $G_e/\omega C_c$.

The High-frequency Gain. In the high-frequency range of the amplifier the interelectrode capacitance susceptances, and these include effects of wiring and sockets, impose an admittance $\underline{Y}_e$ across the two coupling resistors given by the expression

$$\underline{Y}_e = j\omega\,[C_{gpi} + C_{pki} + C_{gko} + C_{gpo}\,(1 - \underline{A}_o)] \tag{3-16}$$

where the subscript i refers to the input tube and o to the output tube. To a first approximation this admittance is a capacitance susceptance equal to

$$j\omega C_e = j\omega[\,C_{gpi} + C_{pki} + C_{gko} + C_{gpo}\,(1 + A_o)] \tag{3-17}$$

Hence the output admittance for the input tube is approximately

$$G_{ci} + G_{go} + j\omega C_e \tag{3-18}$$

where $C_e = C_{pgi} + C_{pki} + C_{gko} + C_{gpo}\,(1 + A_o)$.

The voltage amplification of the stage is

$$\underline{A}_H = \frac{\underline{E}_{go}}{\underline{E}_{gi}} = \frac{-G_{pg}}{G_c + G_{go} + g_{pi} + j\omega C_e} \tag{3-19}$$

In terms of the medium frequency gain the high frequency gain becomes

$$\underline{A}_H = \frac{\underline{A}_M}{1 + j\,\frac{\omega C_e}{G_e'}} \tag{3-20}$$

where $G_e' = G_c + G_{go} + g_{pi}$, and

$$A_H = \frac{A_M}{\sqrt{1 + \left(\frac{\omega C_e}{G_e'}\right)^2}} \quad (3\text{-}21)$$

The reduction in gain in decibels from the medium frequency gain is

$$Db_H = 20 \log_{10} \frac{A_M}{A_L}. \quad (3\text{-}22)$$

The phase shift is $\theta_H = \tan^{-1} \frac{\omega C_e}{G_e'}$

θ_H lags the 180° phase displacement. Figure 3-12 shows how the reduction in gain in decibels and the phase shift depend upon $\omega C_e/G_e'$.

General Considerations. From the relation just developed it is seen that high gain at the medium frequencies can be achieved only by using a tube with a high transconductance and a high plate resistance. R_c and R_{go} are then made as high as practical. R_c is limited by the amount of d-c voltage drop that it is economical to provide in the d-c plate supply and R_{go} is limited by the gas and other conditions that exist in the output tube. The reduction in gain at the low frequencies depends largely upon the size of C_c and R_{go} when $G_{go} \leq 0.2(G_c + g_{pi})$. Increasing either or both of these quantities improves the low frequency gain. However, there is a practical limit to the product of $C_c R_{go}$. This practical limit depends largely upon the amount of gas in the output tube, which causes some non-linear negative grid current and thereby causes C_c to be charged positively to destroy some of the negative bias on the output tube when a sudden impulse occurs. Also unless C_c is a good condenser, such as one with mica dielectric, there will be some leakage of current from the d-c source in the plate circuit which will also put a positive potential across R_{go} that will increase either with an increase of C_c or of R_{go}. Practical limits for R_{go} for small tubes are from 0.5 to 2 megohms, except for special tubes where higher values of R_{go} may be used. A practical limit of $C_c R_{go}$ for a general purpose audio-amplifier is approximately 0.05. Consequently there is a practical limit to the lowest frequency at which a resistance-capacitance-coupled amplifier will provide a specified gain (in decibels from the medium-frequency gain). This limit does not depend much on the value of the gain at the mid-frequency. To achieve low phase displacement at the low frequencies is more difficult than to satisfy the gain requirements.

Low reduction in gain (in decibels) at the high frequencies and high gain at medium frequencies are incompatible for a given set of tubes. The medium-frequency gain can be made high and the reduction in decibels at the high frequencies can be kept low only by the choice of an input tube that has high values of transconductance and plate resistance and an output tube that has a small interelectrode capacitance. For this reason pentodes will give better gain versus frequency characteristics than will triodes, i.e. higher gain for a given decibel reduction at some specified high frequency.

Low-frequency Compensation. Both the gain versus frequency characteristics and the phase shift of a resistance-capacitance-coupled amplifier can be improved in the low-frequency range by placing an additional resistance in series with the coupling resistance and shunting this resistance with a condenser, as shown in Fig. 3-13. The analysis of this circuit is simplified by the use of certain approximations. First assume that R_s is large compared to $1/\omega C_s$ so that $\omega^2 C_s^2 R_s^2 >> 1$. This as-

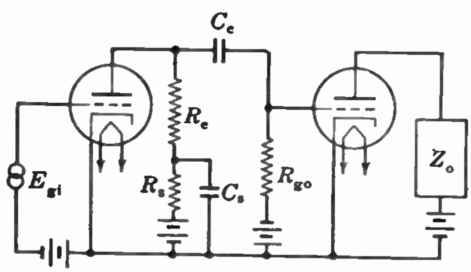


FIG. 3-13. Resistance-capacitance-coupled amplifier with low-frequency compensation.

sumption is justified from practice so long as ω is not too low. Next assume that R_{go} is large compared to R_c so that the only appreciable effect of the C_c and R_{go} circuit is to cause a potential drop and phase shift from the plate of the input to the grid of the output tube. With these assumptions the gain of the stage at the low frequencies becomes

$$\frac{\mathbf{E}_{go}}{\mathbf{E}_{gi}} = \mathbf{A}_L \cong \frac{-\mu\left[R'^2 + \left(\frac{1}{\omega C_s}\right)^2\right]R_{go}}{R_{go}R'(r_{pi} + R') + \frac{r_{pi}}{\omega^2 C_s C_c} + j\left(\frac{r_{pi}R_{go}}{\omega C_s} - \frac{R'(r_{pi} + R')}{\omega C_c}\right)} \qquad (3\text{-}23)$$

where

$$R' = R_c + \frac{1}{\omega^2 C_s^2 R_s}$$

When $R' << r_{pi}$, as for a pentode,

$$\mathbf{A}_L \cong \frac{-\mu\left[R'^2 + \left(\frac{1}{\omega C_s}\right)^2\right]R_{go}}{r_{pi}\left[R_{go}R' + \frac{1}{\omega^2 C_s C_c} + j\left(\frac{R_{go}}{\omega C_s} - \frac{R'}{\omega C_c}\right)\right]} \qquad (3\text{-}24)$$

The approximate expression for the phase shift is

$$\theta_L \cong \tan^{-1} \frac{R_{go}\omega C_c - R'\omega C_s}{R_{go}R'\omega^2 C_s C_c + 1} \tag{3-25}$$

If it were not necessary to pass d-c in the plate circuit R_s could be made infinite and then the gain would be constant with frequency and equal to the medium frequency gain, i.e. $A_L = g_{gpi}R_c$ when the phase shift is completely compensated by making $R_{go}C_c = R_cC_s$. This statement is true only when R_{go} is several times R_c.

When C_s is infinite, the gain and the phase shift vary with the frequency in the usual manner. With C_s finite and of such a value as to satisfy the simplifying assumptions, the phase shift is reduced and can be made almost zero by making $\frac{1}{\omega C_s} R_{go} = \frac{1}{\omega C_c} R'$ or approximately $C_cR_{go} = C_sR_c$ because, since $R' = R_c + \frac{1}{\omega^2 C_s{}^2 R_s}$ and $R_s > \frac{1}{\omega C_s}$ then $\frac{1}{\omega^2 C_s{}^2 R_s} < R_c$. However, as the frequency is decreased the simplifying assumption breaks down and there will be a phase shift. Also, a close examination of expression (3-24) will show that the gain is greater when the circuit is compensated and will be equal to or a little greater than the gain at the medium frequencies when phase compensation is made nearly complete. Expression (3-24) for the low frequency gain of an amplifier gives a fairly accurate analysis only when the grid of the input tube has a fixed bias. When self-bias resistor with a by-pass condenser is used there is some cathode-circuit feed-back effect which reduces the low-frequency gain and shifts the phase in the same direction as that of the shift caused by the coupling condenser C_c. This is shown in Sec. 3-13 where the effect of a self-bias impedance is analyzed. Consequently the design of a low-frequency compensating circuit should take into consideration the effect of the self-bias impedance, which is composed of a resistance and capacitance in parallel.

The medium frequency gain is approximately

$$\underline{A_M} \cong \frac{-\mu R_c}{r_{pi} + R_c} \tag{3-26}$$

when $R_{go} \geqq 10\, R_c$.

3-7. The Inductance-Capacitance-Coupled Amplifier.- Two types of inductance-capacitance-coupled amplifiers are illustrated in Fig. 3-14. In these amplifiers there is very little d-c voltage drop in the plate circuit and so the d-c plate voltage is nearly equal to the supply voltage. The low-frequency gain and phase shift versus frequency of type (a) are worse than for a corresponding resistance-capacitance-coupled

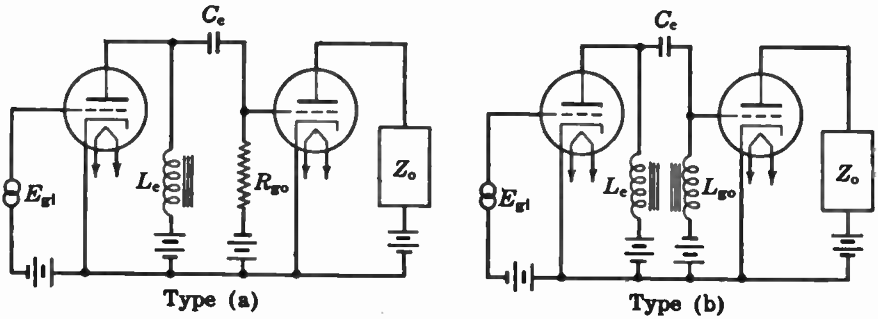


FIG. 3-14. Inductance-capacitance-coupled amplifiers. (a) Single inductance; (b) Double inductance.

amplifier. The impedance of L_c drops with decreasing frequency, and both the gain and the phase shift from grid to plate circuits of the input tube are impaired. In addition, there is a drop in voltage and in phase shift from the plate of the input tube to the grid of the output tube caused by the coupling condenser C_c. The phase shift caused by this coupling condenser C_c and the inductance L_c are in the same direction. Hence the gain and phase shift versus frequency characteristics are poorer than they are for a corresponding resistance-capacitance-coupled amplifier.

The medium-frequency gain of this type of amplifier can be made higher than that of a corresponding resistance-capacitance-coupled amplifier, because the impedance to the input tube is due only to R_{go} plus an almost equally high resistance in parallel that arises from the core loss in the iron core. When R_{go} is adjusted to give the same medium-frequency gain as that obtained from resistance-capacitance coupling, with the same tubes, the high-frequency gain and phase shift characteristics of the two types of amplifiers would be the same because the departure from the medium-frequency gain is due to the same causes; i.e., the tube capacitances place a capacitance in shunt with the coupling impedances. To have reasonably good characteristics at the low frequencies would require a grid resistor lower than that used in the corresponding resistance-capacitance-coupled amplifier and would therefore necessitate a larger coupling condenser.

The approximate expression for the low-frequency gain and phase shift are

$$A_L \cong \frac{A_M}{\sqrt{1 + \left(\frac{r_{p1} R_{go}}{r_{p1} + R_{go}}\right)^2 \left[\frac{1}{\omega^2 L_c^2} + \frac{1}{R_{go}^2 r_{p1}^2 \omega^2 C_c^2}\right]}} \qquad (3\text{-}27)$$

$$\theta_L = \tan^{-1} \frac{\dfrac{1}{\omega C_c} + \dfrac{R_{go} r_{pi}}{\omega L_c}}{\omega(R_{go} + r_{pi}) - \dfrac{r_{pi}}{\omega L_c C_c}} \tag{3-27a}$$

The medium-frequency gain is

$$A_M = \frac{g\,g_{pi}}{g_{pi} + G_o} = \frac{\mu_i R_o}{r_{pi} + R_o} \tag{3-28}$$

where $G_o = G_{go}$ + equivalent core loss conductance.

The high-frequency gain is $A_H = \dfrac{A_M}{\sqrt{1 + \left(\dfrac{\omega C_e}{G_e}\right)^2}}$ (3-29)

where $G_e = g_{pi} + G_{go}$ plus the core loss conductance of L_c, and C_e is the same as that given under 3-18 for the resistance-capacitance-coupled amplifier.

Double Inductance Type. The double inductance type of Fig. 3-14 has characteristics similar to those of the single inductance type at the medium and high frequencies except that the gain depends upon the core loss equivalent resistances of the two inductances in parallel. For the low frequencies the gain is given by the expression

$$A_L = \frac{A_M}{\sqrt{\left(\dfrac{f_r}{f\,Q}\right)^2 + \left(1 - \dfrac{f_r^2}{f^2}\right)^2}} \tag{3-30}$$

where $f_r = \dfrac{1}{2\pi\sqrt{L_{go} C_c}}$ and $Q = \dfrac{\omega_r L_{go}}{r_{pi} + R_{Lgo}}$

when ωL_c is several times R_{go} and at least three times r_{pi}. The curves of Fig. 3-15 show how the gain in decibels and the phase

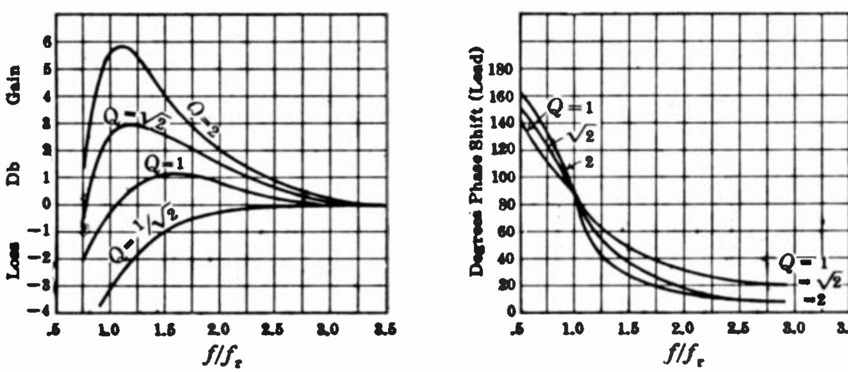


FIG. 3-15. Low-frequency characteristics of a double inductance-capacitance-coupled amplifier. $Q = \dfrac{\omega_r L_{go}}{r_{pi} + R_{Lgo}}$.

shift depart from the medium-frequency characteristics for various values of Q. There is gain compensation at the low frequencies but the phase shift is rather bad for some purposes.

3-8. Transformer Coupling.- Transformer coupling is illustrated in Fig. 3-9. The equivalent circuit of the input stage of the amplifier is shown in Fig. 3-16. This circuit is

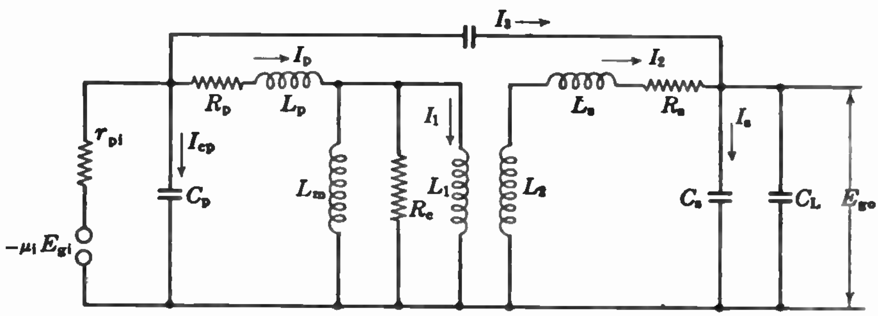


FIG. 3-16. Total equivalent circuit of a transformer-coupled amplifier.

reasonably accurate for those transformers in which the mutual capacitance between the primary and secondary windings can be represented by a single condenser C_m. In the equivalent circuit the symbols have the following significance.

- $-\mu E_{gi}$ is the equivalent voltage generated in the input tube.
- r_{pi} is the plate resistance of the input tube.
- R_p is the resistance of the primary winding.
- R_c is the equivalent resistance of the core losses.
- L_m is the magnetizing inductance. It is the inductance at low frequencies of the primary with the secondary open.
- C_p is the distributed capacitance of the primary winding plus the output capacitance of the input tube.
- L_p is the leakage inductance of the primary winding. This inductance is due to the flux, caused by the load current, which links the primary and does not link the secondary.
- L_s is the leakage inductance of the secondary winding. This inductance is due to the flux, caused by the load current, which links the secondary and does not link with the primary.
- C_s is the distributed capacitance of the secondary.
- C_L is the input capacitance of the output tube.

C_m is the mutual capacitance between the windings.

L_1 and L_2 are fictitious inductances that transfer the load current and voltage. The reactances of L_1 and L_2 are always very high compared to all other reactances and resistances in the circuit. The coupling between L_1 and L_2 is perfect and there are no losses.

N is the ratio of transformation and is equal to L_1/L_2. N is substantially equal to the primary turns divided by the secondary turns. N may be either positive or negative. N is positive if, when the +B and -C terminals of the transformer are joined, a current flowing from the P terminal to the G terminal produces fields by the primary and secondary that are in the same direction in the core.

In the analysis for this type of amplifier leakage inductances for the primary and secondary and the fictitious inductances L_1 and L_2 are used instead of mutual inductance, or coupling coefficient. The reason for this is that the coupling coefficient for a good transformer is 0.95 or better. When the coupling coefficient between two coils is so near unity it is difficult to measure it accurately except by measuring leakage inductance. The total leakage inductance referred to the primary winding is $L_p + N^2L_s$. It is easily measured with good accuracy by shorting the secondary and measuring the primary with an impedance bridge. The inductance so measured is substantially equal to the total leakage inductance referred to the primary. Also with the leakage inductance concept it is easy to see that for the high frequencies there is a series resonant circuit formed by the leakage inductances and the capacitances of the transformer plus the input capacitance of the output tube.

The derivation of the voltage gain of the transformer-coupled stage is carried out by dividing the range of frequencies for which the transformer is intended into the medium frequencies, the low frequencies and the high frequencies. The equivalent circuits for each of these frequency bands are shown in Fig. 3-17. For the medium frequencies the leakage inductive reactances are so small and the capacitance reactance is so high that they have practically no effect upon the performance of the transformer. The magnetizing reactance L_m is so high that the voltage gain is substantially

$$\frac{E_{go}}{E_{gi}} = A_M = \frac{\mu_i R_c}{N(R_c + R_p + r_{pi})}. \qquad (3\text{-}31)$$

For a transformer with a good grade of core material, since the flux density is very low, $R_c >> (R_p + r_{pi})$ and $A_M \cong \mu_i/N$. The phase shift is zero.

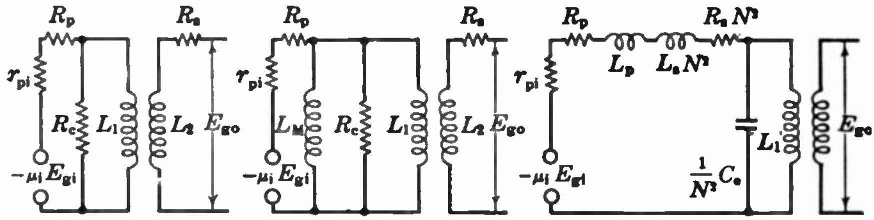


FIG. 3-17. Equivalent circuits of a transformer for medium, low, and high frequencies.

For the low frequencies the effects of the capacitance and leakage inductances are negligible. The low-frequency gain is

$$A_L = \frac{A_M}{\sqrt{1 + \left[\frac{1}{\omega L_m} \frac{(r_{pi} + R_p) R_c}{r_{pi} + R_p + R_c}\right]^2}} = \frac{A_M}{\sqrt{1 + \left(\frac{R_e}{\omega L_m}\right)^2}} \tag{3-32}$$

The phase shift (lead) is

$$\theta_L = \tan^{-1} \frac{(r_{pi} + R_p) R_c}{\omega L_m (r_{pi} + R_p + R_c)}. \tag{3-33}$$

The simplification that results when $R_c >> (r_{pi} + R_p)$ is easily seen in the expression for A_L and θ_L. The low frequency gain can be made to approach A_M and the phase shift approach zero only by making ωL_m several times $\frac{(r_{pi} + R_p) R_c}{r_{pi} + R_p + R_c}$. This is one of the requirements of a good transformer. Approximately, ωL_m should be at least three times $(r_{pi} + R_p)$ for the lowest frequency. The low-frequency characteristics of a transformer-coupled amplifier are very similar to those of a resistance-capacitance-coupled amplifier. The manner in which the gain varies can be obtained from the curve of Fig. 3-12 by using $\frac{r_{pi} + R_p}{\omega L_m}$ in place of $\frac{G_e}{\omega C_e}$.

For the high frequencies only the winding resistance, leakage inductances and capacitance of the transformer affect the gain characteristics. For a very close approximation to the true characteristics the effect of C_p in a well-designed transformer is very small up to the frequency at which the gain of the stage starts falling off rather rapidly. The effect of the mutual capacitance C_m, when C_m is less than 25 per cent of $C_s + C_L$, is largely one of adding C_m to C_s and C_L. For these assumptions the high frequency gain becomes

$$A_H = \frac{\mu_i \frac{1}{\omega C_e} N}{\sqrt{(r_{pi} + R_p + N^2 R_s)^2 + (\omega L_t - \frac{N^2}{\omega C_e})^2}} \tag{3-34}$$

where $L_t = L_p + N^2 L_s$ and $C_e = C_s + C_L + C_m$.

When C_m is more than 25 per cent of $C_s + C_L$ the gain is given approximately by the above expression times the quantity $1 + \frac{NC_m}{C_e} \frac{f^2}{f_r^2}$ and with the resistance r_{pi} replaced by $r_{pi}\ (1 + \frac{NC_m}{C_e})^2$ in which N may be either positive or negative depending upon the directions of the windings. The expression for A_H can readily be thrown into the following form,

$$A_H = \frac{A_M}{\sqrt{(1 - \frac{f^2}{f_r^2})^2 + \frac{f^2}{f_r^2} \frac{1}{Q_r^2}}} \qquad (3\text{-}35)$$

where $f_r = \frac{1}{2\pi \sqrt{\frac{L_t C_e}{N^2}}}$ and $Q_r = \frac{\omega_r L_t}{r_{pi} + R_p + N^2 R_s}$

The curves of Fig. 3-18 show how the gain, in decibels, departs from the medium-frequency gain for various values of Q_r.

The phase shift at the high frequencies is given by

$$\theta_H = 90^\circ + \tan^{-1} Q_r \left(\frac{f}{f_r} - \frac{f_r}{f}\right) \qquad (3\text{-}36)$$

The phase shift curves are also shown in Fig. 3-18.

If the gain of the stage is to remain substantially constant up to a frequency f_h then the transformer must have an f_r higher than f_h or a $Q_r = 1$ when f_r is equal to f_h. Hence f_r

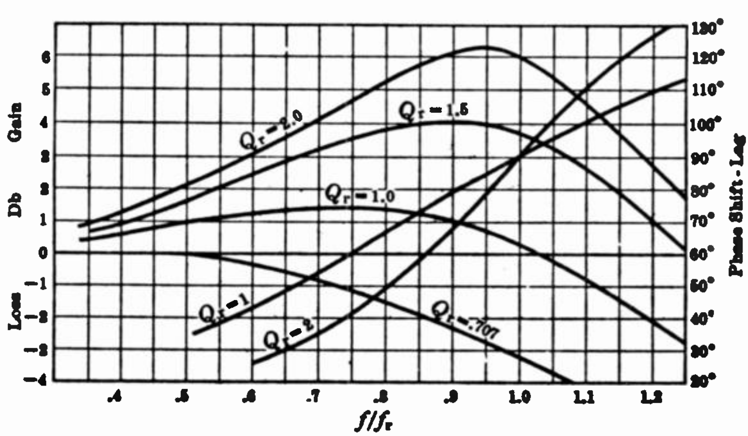


FIG. 3-18. High-frequency characteristics of transformer-coupled amplifier.

should never be less than f_h for a good transformer. When phase shift is a consideration f_r must be quite a lot higher than f_h because the phase shifts badly around f_r.

3-9. Push-Pull and Phase-Inverter Amplifiers.- Voltage amplifiers are often operated in push-pull like the output stage of Fig. 3-19. The advantage of push-pull operation over single-ended operation is the reduction in even harmonic distortion,

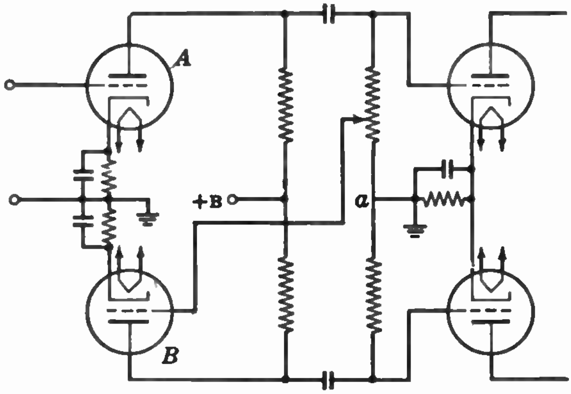


FIG. 3-19. Phase inverter type of push-pull amplifier.

and for transformer-coupled amplifiers there is no d-c flux in the core when the d-c currents of the two tubes are the same. This makes it possible to have a higher primary impedance for the transformer and thereby have better gain-versus-frequency characteristics for the low frequencies as compared to single-ended operation. Filtering of the d-c plate supply can be inferior, decoupling need not be so good and, when self-bias common to both tubes is used, not as much shunt capacitance--theoretically none--is required.

The design of resistance-capacitance-coupled amplifiers is very similar to that of single-ended amplifiers. For the transformer-coupled type the principal difference is in the design of the transformer and this difference is mainly for the high frequencies. If the advantages of push-pull operation are to be achieved over the complete range of frequencies, the transformer must not give rise to secondary voltages that shift in phase and become unequal in magnitude in the high-frequency part of the range. Hence more care must be taken in the design of a push-pull transformer. This also applies to a transformer with single-ended primary and push-pull secondaries.

Phase Inverters. In order to couple a single-ended amplifier to a push-pull amplifier, there must be developed in the coupling circuit between the two amplifiers two voltages 180° out of phase and equal in magnitude. This can be accomplished by a transformer with a single-ended primary and push-pull secondaries. When it is desirable to have resistance-capacitance coupling throughout the amplifier, an additional tube known as a phase inverter is used to develop a voltage 180° out of phase with the output voltage of the single-ended source or section of the amplifier. This is accomplished by tube B of Fig. 3-19. The grid voltage for tube B is obtained from the output voltage of tube A. The output voltage of tube A is 180° out of phase with its own grid voltage over most of the useful frequency range of the amplifier. Hence the grid voltage of tube B is 180° out of phase with the grid voltage of tube A. The grid voltage for tube B can also be derived from the plate coupling resistor of the upper input tube. In either case the resistance between the point of pickoff and neutral, or ground, must equal the total resistance divided by the voltage amplification of the tube A.

A self-balancing type of phase inverter results when a resistor equal to 0.1 to 0.5 of the grid coupling resistor of the output tube is placed between point a and ground and the grid of tube B is connected to point a. When the voltage for the phase inverter tube is obtained in this manner the voltage gain for tubes A and B can be changed quite a little without much unbalance between the grid voltages for the two output tubes.

3-10. Direct-Coupled Voltage Amplifiers.- A direct-coupled amplifier is one that has only resistances for coupling elements between the plate circuit of the input tube and the grid of the output tube. Figure 3-20 illustrates the simplest

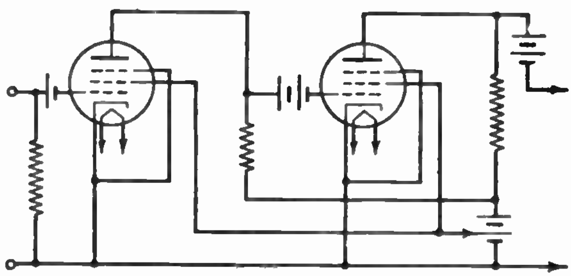


Fig. 3-20. Simplest type of direct-coupled amplifier.

type of direct coupling. Voltage amplifiers of this class are characterized by their low-frequency characteristics. They amplify d-c voltages, and very low-frequency voltages, without phase shift. Their high-frequency characteristics are no better than those of a resistance-capacitance-coupled amplifier because of the shunting effect of the interelectrode capacitances of the tubes.

The amplifier of Fig. 3-20 is not very practical for amplifying changes in d-c applied to the grid of the input tube because of instability due to the batteries. A small change in the potential of a grid or plate circuit battery will cause false changes in the potential across the output resistance. Also because of the separate batteries required the set-up is quite expensive to maintain. The circuit of Fig. 3-21, which is

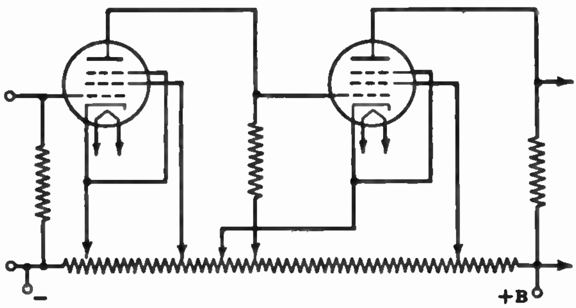


FIG. 3-21. Direct-coupled amplifier with a single d-c supply.

known as the Loftin-White circuit and is somewhat similar to that of Fig. 3-20, avoids the use of more than one battery and may be operated from a medium-high-voltage rectifier that is well filtered and regulated. The current in the voltage divider must be considerably larger than the currents taken by the tubes, so that a change in the tube currents will not disturb the bias, plate, and screen potentials. The total d-c supply voltage is

$$E_{ci} + E_{bi} + E_{co} + E_{bo} + I_{bo}R_o$$

where E_{bi} and E_{bo} are the desired d-c operating plate potentials of the input tube and output tube and E_{ci} and E_{co} are the desired bias potentials of the two tubes. I_{bo} is the drop in potential across the output resistor R_o. Since there is a large difference of potential between the two cathodes it is good practice to have a separate winding on the filament supply transformer for each filament, and to have them well insulated from each other. The gain of the input stage of the amplifier for a change in d-c potential, Δe_{gi}, on the grid of the input tube is

$$\frac{\Delta e_{go}}{\Delta e_{gi}} = \frac{-\mu R_c}{R_c + r_p} R_c$$

where Δe_{gi} is the change in the input voltage and Δe_{go} is the resulting change in voltage across the grid of the output tube. The response of the amplifier to an a-c voltage is independent of the frequency up to the higher frequencies where the inter-electrode capacitances of the tubes place a reactance in shunt with the coupling resistor comparable with the magnitude of the coupling resistor. Negative feed-back may be applied to the amplifier by placing a resistor in the cathode circuit; it will improve the stability.

Another direct-coupled amplifier that has a gain nearly equal to the amplification factor of the tube provided it is used with a very high load impedance has been proposed by Schmitt[7] for a-c applications. This amplifier, shown in Fig. 3-22, makes use of a pentode for a resistance through which the

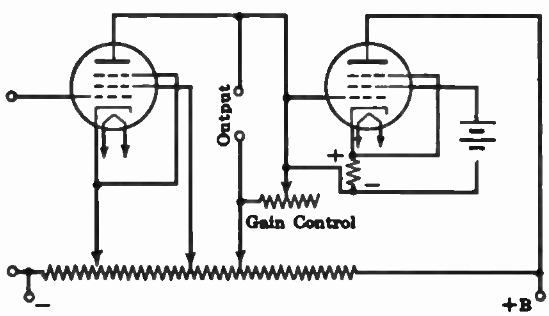


Fig. 3-22. High-gain direct-coupled amplifier.

plate of the amplifier tube receives its d-c voltage. The pentode, when operated on the nearly flat portion of its I_b vs E_b characteristic has an a-c plate resistance several times its d-c resistance. The d-c voltage drop through the tube is furnished by an addition to the d-c plate supply. With some of the modern pentodes, having amplification factors as high as 6000 with a plate resistance of 0.75 megohms, a very high gain (about 2000) can be obtained with a d-c supply of 500 volts.

There are many other types of direct-coupled amplifiers. Some employ degenerative feed-back which greatly improves their stability. Others use tubes in push-pull with phase inverters for single-ended input. Push-pull should be used wherever possible because it helps to improve stability by reducing some of the effect from changes in d-c supply potentials.

7. Schmitt, Otto H.A., "A Method of Realizing the Full Amplification of High-mu Tubes." Rev. Sci. Inst. Dec. 1933.

3-11. Amplifiers With Degenerative (Negative) Feed-back.- A degenerative feed-back amplifier is one in which a voltage from the output is introduced into the input of the amplifier in such a manner that the net input grid voltage is reduced, thus resulting in a reduction of the gain of the amplifier. The reduction in gain is a disadvantage that is offset by several advantages such as a reduction of distortion, an improved frequency characteristic, and a higher degree of stability, all of which are inherent in the feed-back amplifier. In many applications of multi-stage amplifiers the aim is to produce a given voltage or power output without exceeding a certain allowable amount of distortion of either wave form, phase or frequency. Often the requirements as to distortion are so severe that extra expense and complexity of construction are warranted if the desired results are achieved. Hence, if degenerative feed-back will improve the characteristics of the amplifier its application may be well worth the extra expense even though more stages will be required to give the desired amount of amplification.

Simple Type of Feed-back. The simplest type of degenerative feed-back is shown in Fig. 3-23. In this circuit the feed-back voltage is introduced into the input circuit between the input voltage E_1 and cathode and is designated E_f. When the tube is operated so that there is no grid current the voltage E_f is directly proportional to the output voltage E_o. Let the voltage amplification A be the actual voltage amplification of the tube, i.e., $A = E_o/E_g$. Assume the capacitance C, which is used to keep the d-c out of the feed-back circuit, is so large that the voltage drop and phase shift due to C are negligible and further that R is large compared to R_o, that is, the output impedance of the tube is substantially R_o. Then

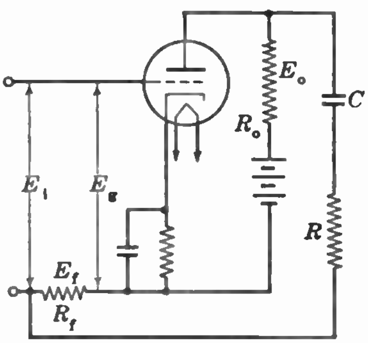


FIG. 3-23. Simple amplifier circuit illustrating the basic principle of feed-back.

$E_f = E_o \frac{R_f}{R_f + R}$. Subtracting both sides from E_g gives

$$E_g - E_f = E_g - E_o \frac{R_f}{R_f + R}$$

But $$E_g - E_f = E_1 \text{ or } E_g = E_1 + E_f$$

Hence $$E_1 = E_g - E_o \frac{R_f}{R_f + R}$$

From the relation $\underline{A} = \frac{\underline{E}_o}{\underline{E}_g}, \quad \underline{E}_g = \frac{\underline{E}_o}{\underline{A}}$

Consequently, $\underline{E}_i = \underline{E}_o \left(\frac{1}{\underline{A}} - \frac{R_f}{R_f + R}\right)$

and finally,

$$\underline{A}_{fb} = \frac{\underline{E}_o}{\underline{E}_i} = \frac{\underline{A}}{1 - \underline{A}\dfrac{R}{R_f + R}} \qquad (3\text{-}37)$$

where $\underline{A}$ is the gain of the amplifier without feed-back, but with C, R_f, and R connected in series from plate to cathode. For the simple amplifier shown $\underline{A}$ has a negative real value equal to A. Hence

$$A_{fb} = \frac{E_o}{E_i} = \frac{A}{1 + A\dfrac{R}{R_f + R}}$$

Thus it is seen that the feed-back effect is such as to reduce the voltage gain. This is a disadvantage of degenerative feed-back. However, the important advantages above referred to will be discussed below.

<u>General Expression for Feed-back</u>. Since degenerative feed-back may be applied in several different ways it is better to have a more general expression than the one given above. In the above expression the quantity $\underline{A}\, R/(R_f + R)$ is called the feed-back factor. In the more general discussion the feed-back factor will be designated as $\underline{A}\,\underline{\beta}$. Hence the general expression for an amplifier that has degenerative feed-back is

$$\underline{A}_{fb} = \frac{\underline{A}}{1 - \underline{A}\,\underline{\beta}} \qquad (3\text{-}38)$$

where $\underline{A}_{fb}$ is the voltage amplification with feed-back, $\underline{A}$ is the vector voltage amplification that the portion of the amplifier which is controlled by feed-back would have without feed-back. $\underline{\beta}$ is the vector ratio of the feed-back voltage to the voltage which exists at the higher level point at which $\underline{A}$ is reckoned. The exact expression for $\underline{\beta}$ for the circuit of Fig. 3-23 is $\underline{\beta} = R_f/(R + R_f - j\frac{1}{\omega C})$. When $\frac{1}{\omega C}$ is very small compared to R, $\underline{\beta}$ is nearly equal to $R_f/(R + R_f)$. For degenerative feed-back $|1 - \underline{A}\,\underline{\beta}|$ must be <A. This is one limitation which determines the manner in which degenerative feed-back may be applied in a single or multi-stage amplifier; it means that the feed-back voltage to the grid of any stage of an amplifier must be more than 90° out of phase with the input voltage to that particular stage.

When the feed-back factor $|\underline{A}\,\underline{\beta}|$ is large compared to 1 the gain of the amplifier becomes

$$\text{Gain of amplifier} = -\frac{1}{\underline{\beta}} \quad (A\,\beta \gg 1) \qquad (3\text{-}39)$$

Under these conditions the gain of the amplifier depends only upon the feed-back circuit. Hence any change in $\underline{A}$ which does not destroy the relation $A\beta \gg 1$ does not produce any change in the gain of the amplifier. Furthermore the gain and phase shift versus frequency characteristics are established entirely by the feed-back circuit.

To make $|\underline{A}\,\underline{\beta}|$ large compared to 1 would require β to be large compared to 1/A. Hence under these conditions the gain of the amplifier will be only a fraction of the gain of the same amplifier without feed-back. The input voltage $\underline{E}_i$ to the amplifier will have to be raised to a value $\underline{A}\,\underline{\beta}$ times the voltage necessary without feed-back in order to produce the same output voltage $\underline{E}_o$ with feed-back. From the above expression for the gain of the amplifier it is readily seen that the input voltage is given by the expression $\underline{E}_i = -\underline{\beta}\,\underline{E}_o$.

Effect of Feed-back on Stability. Returning to the general expression for the gain of the amplifier with feedback, namely,

$$\underline{A}_{fb} = \frac{\underline{A}}{1 - \underline{A}\,\underline{\beta}}$$

it is apparent upon examination of this expression that any amount of feed-back will cause the gain of the amplifier to be changed less than any given change in $\underline{A}$. $\underline{A}$ may change for several reasons: change of tubes, change in operating voltage, change in load resistance. Without feed-back a change of 10 per cent in A causes a change of 10 per cent in the gain of the amplifier. With feed-back such that $A\beta = -1$ a 10 per cent increase in A causes only a 4.76 per cent increase in the gain of the amplifier. The curves of Fig. 3-24 show how the percentage change in gain of the amplifier depends upon A for various percentage changes in A. The upper curve is for a 20 per cent increase in A and the lower for a 10 per cent increase. The curves of Fig. 3-25 show how the normal gain of the amplifier depends upon A. By normal is meant the gain for any standard set of conditions.

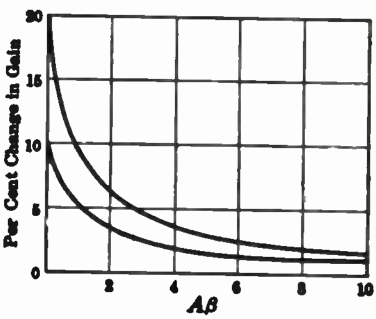


FIG. 3-24. Curves showing how feed-back affects the stability of an amplifier. Angle of A = 180°.

Effect of Feed-back on Phase Shift. Feed-back also reduces phase shift in an amplifier. That is, it brings the input and output voltages more nearly 180° out of phase. This is shown by the vector diagrams of Fig. 3-26. Assume the output impedance of the circuit of Fig. 3-23 is composed of resistance and inductance, and further assume $1/\omega C \ll R$. Then with no feed-back, i.e., $A\beta = 0$ the vector diagram of Fig. 3-26a

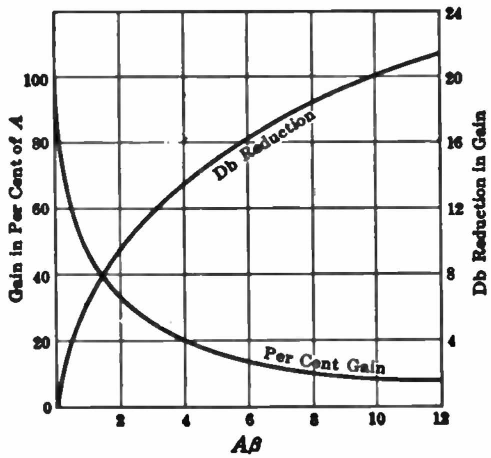
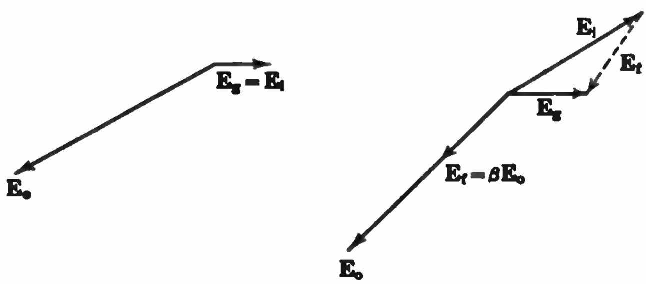


FIG. 3-25 Curves showing how feed-back affects the gain of an amplifier. Angle of A = 180°.

applies, and $\underline{E}_o$ and $\underline{E}_i$ are less than 180° out of phase. With feed-back Fig. 3-26b applies. $\underline{E}_f$ is the voltage fed back into the input circuit. $\underline{E}_i$ and $\underline{E}_o$ are nearer to 180° out of phase



(a) Load impedance inductive $A\beta = O$.

(b) Load impedance inductive $A\beta \neq O$. E_o and E_g same as for figure (a); $E_f = \beta E_o$; $\beta = R/(R_f + R)$.

FIG. 3-26. Effect of feed-back on phase shift.

than for the case of no feed-back. Hence in a feed-back amplifier the change of phase of the amplification with frequency will be less than in the same amplifier without feed-back provided the feed-back circuit is free of phase shift, or has phase shift of a compensating nature.

The general expression which shows how phase shift is reduced is derived as follows: First assume $\underline{\beta}$ real and the feed-back voltage in phase with the output voltage. Second, assume $180^\circ + \theta$ is the phase angle of $\underline{A}$. Then vector $\underline{A}$ may be written as $\underline{A} = -A\ (\cos\theta + j\sin\theta)$. Hence

$$\underline{A}_{fb} = \frac{-A\ (\cos\theta + j\sin\theta)}{1 + A\beta\ (\cos\theta + j\sin\theta)} \tag{3-40}$$

Now let $$1 + A\beta(\cos\theta + j\sin\theta) = U\underline{/\alpha}$$

where $$\alpha = \tan^{-1}\frac{A\beta\sin\theta}{A\beta\cos\theta + 1}$$

and $$U = \sqrt{1 + 2\,A\beta\cos\theta + A^2\beta^2}$$

Therefore the gain of the amplifier is

$$\underline{A}_{fb} = \frac{-A\underline{/\theta}}{U\underline{/\alpha}} = \frac{-A}{U}\underline{/\theta - \alpha} \tag{3-41}$$

As an example suppose $A\beta = 1$ and $\theta = 30^\circ$, i.e., the output voltage of the amplifier without feed-back is $180^\circ + 30^\circ$ out of phase with the input voltage. Under these conditions

$$U = \sqrt{1 + 1.732 + 17} = 1.935$$

and $$\alpha = \tan^{-1}\frac{A\beta\sin\theta}{A\beta\cos\theta + 1} = \tan^{-1}\frac{0.5}{1.866} = \tan^{-1} 0.268$$

or $$\alpha = 15^\circ$$

Finally $$A_{fb} = \frac{-A}{1.935}\underline{/30^\circ - 15^\circ} = \frac{-A}{1.935}\underline{/15^\circ}$$

Similarly when $$A\beta = -2 \text{ and } \theta = 30^\circ$$

$$A_{fb} = \frac{-A}{2.94}\underline{/\text{Appx. } 10^\circ}$$

Effect of Feed-back on Gain versus Frequency. Previously it was stated that feed-back reduces the change in the gain of the amplifier with frequency. This is readily seen when $A\beta \gg 1$ because the gain of the amplifier is equal to $-1/\beta$ and if β is independent of frequency then the gain is independent of frequency. To show how feed-back improves the gain versus frequency characteristics of an amplifier it will be assumed that the load impedance is inductive for the simple single-tube case. Then without feed-back

$$\underline{A} = \frac{-\mu(R_o + jX_o)}{R_o + r_p \pm jX_o} \text{ and } A = \frac{\mu\sqrt{R_o^2 + X_o^2}}{\sqrt{(R_o + r_p)^2 + X_o^2}} \tag{3-42}$$

In order to make the case as severe as possible let $R_o \ll X_o$.

Then $$A = \frac{\mu}{\sqrt{\dfrac{(R_o + r_p)^2}{X_o^2} + 1}} \tag{3-43}$$

Since $X_o = 2\pi f L_o$ the gain will increase with increasing frequency until $(R_o + r_p)/X_o$ is small compared to 1. With feed-back, assuming β is independent of frequency and the feed-back circuit has no appreciable effect upon the output impedance of the tube

$$\underline{A}_{fb} = \frac{\dfrac{-\mu(R_o + jX_o)}{r_p + R_o + jX_o}}{1 + \dfrac{\mu(R_o + jX_o)}{r_p + R_o + jX_o}\beta} \tag{3-44}$$

$$\text{and} \quad A_{fb} = \frac{\mu\sqrt{R_o^2 + X_o^2}}{\sqrt{[R_o(1 + \mu\beta) + r_p]^2 + X_o^2(1 + \mu\beta)^2}}$$

Again when $R_o \ll X_o$

$$A_{fb} = \frac{\mu}{\sqrt{\dfrac{[R_o(1 + \mu\beta) + r_p]^2}{X_o^2} + (1 + \mu\beta)^2}} \tag{3-45}$$

In this case the gain of the amplifier will increase with increasing frequency and reach a value equal to $\dfrac{\mu}{(1 + \mu\beta)}$ when $\dfrac{R_o(1 + \mu\beta) + r_p}{X_o}$ is small compared to $1 + \mu\beta$. For any given amplifier of this type when $[R_o(1 + \mu\beta) + r_p]/X_o$ is large compared to $(1 + \mu\beta)$ or $(R_o + r_p)/X_o$ is large compared to 1 the gain is practically the same whether feed-back is employed or not, but as $[R_o(1 + \mu\beta) + r_p]/X_o$ approaches $1 + \mu\beta$, or $(R_o + r_p)/X_o$ approaches 1, feed-back reduces the gain a substantial amount when $\mu\beta$ is equal to or greater than 1. Hence the gain versus frequency characteristics of the feed-back amplifier are more nearly constant than similar characteristics of the same amplifier without feed-back. However, a value of feed-back which results in a marked improvement in the gain versus frequency characteristics reduces the gain of the amplifier so much that it becomes questionable whether any improvement has resulted when one considers that it is necessary to add more amplifiers to the system in order to bring the gain with feed-back up to the gain without feed-back. This is considered further in the discussion of feed-back in multi-stage amplifiers.

Feed-back Applied to Multi-stage Amplifiers. It is generally difficult to apply the proper kind of feed-back to each of the separate stages of a multi-stage amplifier. It is also sometimes impossible to introduce the right kind of feed-back in any other stage of an amplifier except the first stage unless the stages are coupled together by transformers. Consequently no generalized method for applying feed-back in a multi-stage amplifier can be given and each type of amplifier must be treated separately. Some of the different ways of applying simple and multiple feed-back in multi-stage amplifiers will be discussed later.

Simple feed-back can usually be applied between the output and the input of a multi-stage amplifier whenever the source driving the amplifier can be raised to a potential above the cathode of the first tube as shown in Fig. 3-27. Assume that

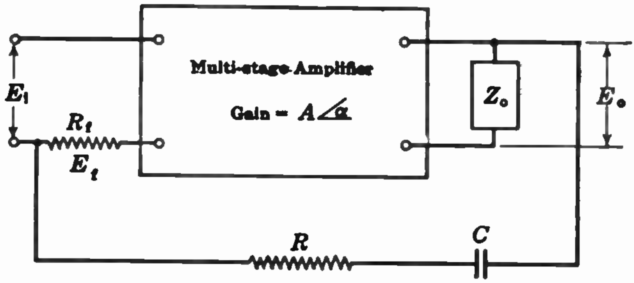


FIG. 3-27. Illustrating feed-back from the output to the input of a multi-stage amplifier.

the amplifier is made up of n stages and let these stages be numbered 0, 1, 2, etc. from the output end. Let $A_1\angle\varphi_1$ be the gain of any stage. Then the total gain of the amplifier without feed-back, but with the feed-back circuit connected across the output impedance, is

$$\underline{A} = A_0 \times A_1 \times A_2 \ldots\ldots \angle\varphi_0 + \varphi_1 + \varphi_2 + \ldots = A_t\angle\varphi \quad (3\text{-}46)$$

where $A_t = A_0 \times A_1 \times A_2 \ldots$ and $\varphi_t = \varphi_0 + \varphi_1 + \varphi_2 + \ldots\ldots$

(Note that each φ is phase displacement and not phase shift.)

$$\underline{A}_{fb} = \frac{A_t\angle\varphi_t}{1 - A_t\beta\angle\varphi_t} = \frac{A_t\angle\varphi_t}{1 - A_t\beta\,(\cos\varphi_t + j\sin\varphi_t)} \quad (3\text{-}47)$$

$$\underline{A}_{fb} = \frac{A_t}{U}\angle\varphi_t - \alpha$$

where
$$U = \sqrt{1 - 2A_t\beta\cos\varphi_t + (A_t\beta)^2}$$

and
$$\alpha = \tan^{-1}\frac{-A_t\beta\sin\varphi_t}{1 - A_t\beta\cos\varphi_t}$$

When $\underline{\beta}$ is a general vector quantity having a value $\underline{\beta} = \beta\angle\psi$

then
$$\underline{A}_{fb} = \frac{A_t}{U}\angle\varphi_t - \alpha_t'$$

where
$$U = \sqrt{1 - 2A_t\beta\cos(\varphi_t + \psi) + (A_t\beta)^2}$$

and
$$\alpha_t' = \tan^{-1}\frac{-A_t\beta\sin(\varphi_t + \psi)}{1 - A_t\beta\cos(\varphi_t + \psi)}$$

In the above expressions φ_t is the phase displacement between the output voltage and input voltage of the amplifier without feed-back and A_t is the total voltage amplification without feed-back. With feed-back the gain of the amplifier is $\frac{A_t}{U}$ and the phase displacement between output and input voltage is $\varphi_t - \alpha_t$.

The feed-back will be degenerative when cos $(\varphi_t + \psi)$ has a negative value. This means the output voltage must be from 90° to 270° out of phase with the input voltage when the phase displacement in the feed-back circuit is zero. Hence in a resistance-capacitance-coupled amplifier for the way in which feed-back is applied in Fig. 3-27 there must be an odd number of stages in the amplifier. In a transformer-coupled amplifier there are no such restrictions on the number of stages because the transformers can be wound to give the correct phase displacement for either an even or an odd number of stages. There are, however, other restrictions in any type of amplifier which must be considered. Oscillations will result when the denominator of (3-47) becomes zero, i.e.,

$$1 - A_t\beta \cos(\varphi_t + \psi) + j\, A_t\beta \sin(\varphi_t + \psi) = 0.$$

Hence feed-back must be applied in such a way that oscillations cannot occur and further such that if there be regeneration in the desired frequency band of the amplifier it does not produce undesired gain versus frequency characteristics.

Therefore when attempting to design an amplifier of any type that is to include degenerative feed-back one must be certain of the voltage amplification and phase shift characteristics of the amplifier without feed-back not only over the band of frequencies for which the amplifier is intended but also frequencies below and above this band. In many cases it may not matter what the gain versus frequency characteristics of the ultimate amplifier are outside the band of interest so long as sustained oscillations do not take place. When the feed-back circuit includes transformers special care must be exercised. Referring to the theory of the transformer-coupled amplifier it will be noted that the phase shift of a single stage is equal to 90° at f_r and exceeds 90° for frequencies above the resonant frequency. When the transformer has a Q_r greater than unity, the voltage amplification for frequencies near the resonant frequency is greater than it is at the mid-frequency. For frequencies above f_r, the amplification drops off so rapidly that trouble from regeneration in a single stage amplifier is not usually encountered. With more than one stage trouble from regeneration is usually the rule unless special care is taken to attenuate the feed-back voltage. The presence of the feed-back resistance across the secondary of the transformer modifies these statements somewhat*.

<u>Reduction of Amplitude Distortion and Other Disturbances by Feed-back</u>. Let the total output voltage of an amplifier with

feed-back be written as follows:

$$e_o' = Ae_i' + e_d$$

where e_d is the distortion or disturbing voltage that would exist if there were no feed-back and e_i' is the input voltage from the source to be amplified plus the feed-back voltage, or

$$e_i' = e_i + \beta(Ae_i' + e_d)$$

Then
$$e_i' = \frac{e_i}{1 - A\beta} + \frac{\beta e_d}{1 - A\beta}$$

and finally, by substituting the value of e_i' in the expression for e_o' there results

$$e_o' = \frac{Ae_i}{1 - A\beta} + \frac{e_d}{1 - A\beta} \tag{3-49}$$

Hence the distortion or disturbing voltage e_d is reduced in the same manner as the gain of the amplifier. For the case of non-linear distortion the reduction shown is true only when the output voltage is the same for feed-back as it is without feed-back and the generated distortion voltage e_d is not affected by feed-back.

Output Terminal Impedance of a Feed-back Amplifier. There are two cases for consideration of the effect of feed-back on the output impedance of the amplifier: (1) when the feed-back is voltage controlled, and (2) when the feed-back is current controlled. If (see Fig. 3-27) the input terminals of an amplifier are shorted and an a-c voltage $\underline{E}_o$ is applied to the output terminals with the load impedance removed the current from the source will be $\underline{I}_o = \underline{E}_o/\underline{Z}_o$ where $\underline{Z}_o$ is the output terminal impedance, when there is no feed-back.

With voltage controlled feed-back the current will be $\underline{I}_o' = \dfrac{\underline{E}_o - \underline{A}\beta\underline{E}_o}{\underline{Z}_o}$ where $\underline{A}$ is the voltage gain without feed-back from the input to the output terminals, and $\underline{A}\,\underline{\beta}\,\underline{E}_o$ is the voltage fed back and amplified to the output. Hence

$$\underline{Z}_o' = \frac{\underline{E}_o'}{\underline{I}_o'} = \frac{\underline{Z}_o}{1 - \underline{A}\,\underline{\beta}} \tag{3-50}$$

where $\underline{Z}_o'$ is the apparent impedance measured across the output terminals of an amplifier with voltage controlled feed-back. Hence, since $|1 - \underline{A}\,\underline{\beta}|$ must be greater than unity for negative feed-back the output impedance of an amplifier with voltage controlled feed-back is lowered.

With current controlled feed-back, where the feed-back voltage is derived from a series resistor R and the resistance is small compared to the output impedance, the expression for the output impedance is obtained from

$$I_o'' = \frac{E_o + A\, I_o''\, R}{Z_o}$$

The solution for Z_o'' gives

$$Z_o'' = Z_o - A\, R$$

where A is the voltage gain of the amplifier from the input to the output terminals without feed-back and Z_o'' is the apparent output impedance measured across the output terminals. Hence for current controlled feed-back the apparent output impedance of the amplifier is raised because A has a negative real value for degenerative feed-back.

Degenerative Feed-back Circuits and Considerations. There are many ways in which degenerative feed-back can be applied in single and multi-stage amplifiers. In all cases degenerative feed-back either stabilizes the current or the voltage of that part of the amplifier from which it is derived, i.e. the feed-back is either current controlled or voltage controlled. The circuits of Figs. 3-23 and 3-27 show voltage controlled degenerative feed-back.

When the shunt capacitance of a self-bias resistor in an amplifier is removed from all or a part of the resistor the feed-back is controlled by the plate current of the tube and the plate current is therefore stabilized. This is illustrated in Fig. 3-28. The effect of feed-back of the kind shown in this figure is different from that shown in Fig. 3-23, when Z_o is not a pure resistance. When Z_o is a pure resistance there is no difference between current and voltage controlled feed-back. When Z_o is a coupling circuit for a resistance-capacitance-coupled amplifier plate current controlled feed-back gives no improvement in the voltage gain versus frequency characteristics of the stage, but does reduce distortion and improve stability. When the feed-back is controlled by the voltage across R_g, there is also an improvement in the voltage gain versus frequency characteristics. Consequently in order to stabilize the output voltage of an amplifier the feed-back voltage must be proportional to the output voltage.

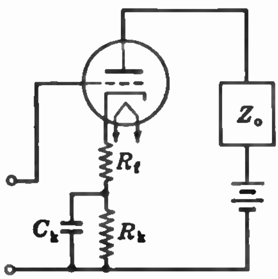


FIG. 3-28. Current controlled feed-back. Feed-back voltage developed across R_f.

Feed-back voltage derived as shown in Fig. 3-23 cannot be introduced into any stage of the amplifier, except the input, unless there is a transformer coupling immediately preceding the point of feed-back. In other than transformer-coupled amplifiers, cathode circuit feed-back illustrated in Fig. 3-30 is generally used. The overall feed-back embraces both tubes and

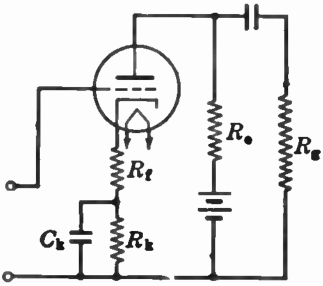


Fig. 3-29. Cathode resistor feed-back in a stage of resistance-capacitance coupling.

the output impedance and in addition there is some plate current feed-back in the first tube. This is called multiple feed-back.

It was shown above that feed-back controlled by the load current in an amplifier raises the apparent output impedance of an amplifier. Consequently the load impedance can be changed a reasonable amount without much effect upon the load current. Use is sometimes made of this property when an amplifier is operated with load impedances which are changed from time to time. On the other hand if the load voltage is to remain nearly constant with changes in load impedance feed-back controlled by the load voltage is used.

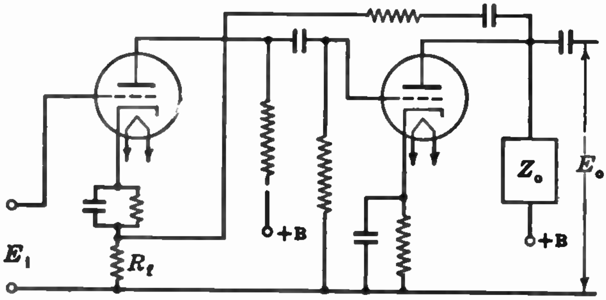


Fig. 3-30. Cathode circuit feed-back in a two-stage resistance-capacitance-coupled amplifier.

There are many interesting ways of applying degenerative feed-back in single-ended and push-pull amplifiers. The student, with the above fundamentals in mind, can understand, and even devise circuits appropriate to, the various applications of feed-back found in more extensive treatments in the reference material.

3-12. Feed-back in a Multi-stage Amplifier from Common Impedance in the D-C Source; Decoupling Circuits.- In a multi-stage amplifier the plate circuits are usually energized from the same d-c source. This d-c source may be either a rectifier or batteries. Either type of d-c source usually has sufficient impedance or resistance to cause an alternating voltage to be set up, principally by the current from the last tube, which is impressed in the plate circuits of all the other tubes. Let this common impedance coupling be symbolized by Z_c as shown in Fig. 3-31.

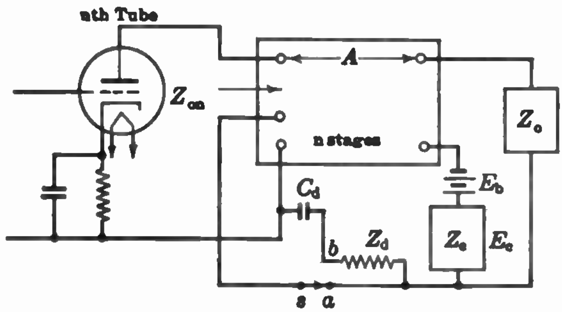


FIG. 3-31. Diagram illustrating common impedance coupling between output stage and nth stage from output.

In a multi-stage amplifier there will be regenerative, or degenerative, effects between the output of the amplifier and each of the stages because of the common impedance couplings. A treatment which would take into account all of these effects would be very complicated in the final results. The worst effect is between the plate circuits of the first and last tubes of the amplifier. Since the object of this analysis is to bring out a way in which the regenerative effect can be reduced to such an extent that the amplifier will perform properly it is necessary to treat only the worst effect and then apply the same cure to the other stages if they need it.

Referring to Fig. 3-31 let n be the number of stages between the input tube and the output impedance $\underline{Z}_o$ of the output tube; let $\underline{Z}_{on}$ be the output impedance to the nth tube of the amplifier as shown in Fig. 3-31. Let $\underline{A}$ be the voltage amplification from the plate circuit of the input tube to the plate circuit of the output tube. Assume the voltage which is set up in the common impedance $\underline{Z}_c$ by the plate current of the output tube is $\underline{E}_c = -\underline{I}_{po}\underline{Z}_c$. For all practical purposes $\underline{I}_{po} = \frac{-\underline{E}_o}{\underline{Z}_o}$ because Z_c is generally small compared to Z_o and does not have much effect upon the output impedance of the amplifier. Consequently

$$\underline{E}_c = \frac{-\underline{Z}_c\underline{E}_o}{\underline{Z}_o}.$$

The plate voltage of the nth tube from the output end of the amplifier is

$$\underline{E}_{pn} = \underline{E}_c - (\underline{E}_{gn} g_{gp} + \underline{E}_{pn} g_p)\underline{Z}_{on}$$

because Z_c is very small compared to Z_{on}.

Also $$\underline{E}_{pn}\underline{A} = \underline{E}_o \text{ or } \underline{E}_{pn} = \frac{\underline{E}_o}{\underline{A}},$$

where $\underline{A}$ is the voltage amplification from the plate of the nth tube to the output impedance $\underline{Z}_o$.

Hence $$\frac{\underline{E}_o}{\underline{A}} = \frac{\underline{Z}_c \underline{E}_o}{\underline{Z}_o} - (\underline{E}_{gn} g_{gp} + \frac{\underline{E}_o}{\underline{A}} g_p)\underline{Z}_{on}$$

or $$\underline{E}_o(1 + \underline{Z}_{on} g_{pn} - \frac{\underline{Z}_c}{\underline{Z}_o} A) = -\underline{E}_{gn}\underline{Z}_{on} g_{gpn}\underline{A},$$

and finally, $$\frac{\underline{E}_o}{\underline{E}_{gn}} = - \frac{\underline{Z}_{on} g_{gpn}\underline{A}}{1 + \underline{Z}_{on} g_{pn} - \frac{\underline{Z}_c}{\underline{Z}_o}\underline{A}} \quad (3\text{-}52)$$

Since the quantity $-\frac{Z_{on}\mu_n A}{r_{pn} + \underline{Z}_{on}}$ is the overall voltage amplification of the system when there is no common impedance effect or when $\underline{Z}_c = 0$, the overall gain of the amplifier with common impedance coupling effect is

$$\frac{\underline{E}_o}{\underline{E}_{gn}} = \frac{\text{Gain of the amplifier with no common impedance coupling}}{\left[1 - \frac{\underline{Z}_c}{\underline{Z}_o}\frac{A}{1 + \underline{Z}_{on} g_{pn}}\right]} \quad (3\text{-}53)$$

where the quantity $\frac{Z_c A}{\underline{Z}_o(1 + \underline{Z}_{on} g_{pn})}$ is the feed-back effect due to common impedance coupling between the output stage and the input stage of the amplifier. This quantity may have either a positive or negative real term and a positive or negative j term. Hence the effect may be either regenerative or degenerative. In either case, since the quantity depends upon frequency, the gain of the amplifier will vary with frequency in a manner differently than it would if there were no common impedance coupling. When $\frac{\underline{Z}_c}{\underline{Z}_o}\frac{A}{(1 + \underline{Z}_{on} g_{pn})}$ has a real term which is positive and equal to unity for a frequency at which the j term vanishes, the amplifier will generate sustained oscillations at a frequency which will make the j term zero. Under these conditions the system will be entirely useless as an amplifier. This is a very common experience in practice unless means are taken to reduce the feed-back effect from common-impedance coupling.

Decoupling Circuits. The method for reducing, or eliminating the effect of, common impedance coupling is to attenuate the voltage $\underline{E}_f$ between the impedance $\underline{Z}_c$ and the plate circuit of the nth stage and a condenser from the end of the series resistor to cathode or ground of the nth tube. This is called a decoupling circuit and is illustrated in Fig. 3-31 by throwing switch s to position b. This places an impedance $\underline{Z}_d$ in series and a condenser C_d in shunt with the d-c plate supply to the nth tube.

Let the attenuation factor, i.e., the factor by which the voltage $\underline{E}_c$ is reduced for the plate circuit of the nth stage, be represented by $\underline{D}$. Then

$$\frac{\underline{E}_o}{\underline{E}_{gn}} = \frac{\text{Gain with no common impedance coupling}}{1 - \dfrac{\underline{Z}_c}{\underline{Z}_o}\,\dfrac{\underline{A}\,\underline{D}}{(1 + \underline{Z}_{on}\,g_{pn})}} \qquad (3\text{-}54)$$

For the case of a condenser C_d and an impedance $\underline{Z}_d$ the decoupling factor $\underline{D} = \dfrac{j/\omega C_d}{\underline{Z}_d}$ when $\dfrac{1}{\omega C_d} << Z_d$ at the lowest frequency for which the amplifier is designed. In order to make the decoupling circuit effective and to provide a low impedance shunt for the a-c plate current of the nth tube it is necessary that $1/\omega C_d$ be small compared to Z_d.

When $\underline{Z}_c$ is known it is a matter of adjusting C_d and $\underline{Z}_d$ to such values that $\underline{D}$ makes the magnitude of the quantity $\dfrac{\underline{Z}_c}{\underline{Z}_o}\,\dfrac{\underline{A}\,\underline{D}}{(1 + \underline{Z}_{on} g_{pn})}$ small compared to unity for all frequencies. The reactance $1/\omega C_d$ must be small compared to the outout impedance of the nth tube and R_d should not be so large as to materially reduce the d-c voltage to the nth tube unless such a reduction is necessary. When $\underline{Z}_c$ is not known but it is certain that Z_c is small compared to Z_o it will be satisfactory for most cases to make $D = 1/A$. This will result in a value for $\dfrac{\underline{Z}_c}{\underline{Z}_o}\,\dfrac{\underline{A}\,\underline{D}}{1 + \underline{Z}_{on} g_{pn}}$ that is small enough compared to unity to make the amplifier behave in a proper manner.

Figure 3-32 illustrates a three stage resistance-capacitance-coupled amplifier with a two section decoupling

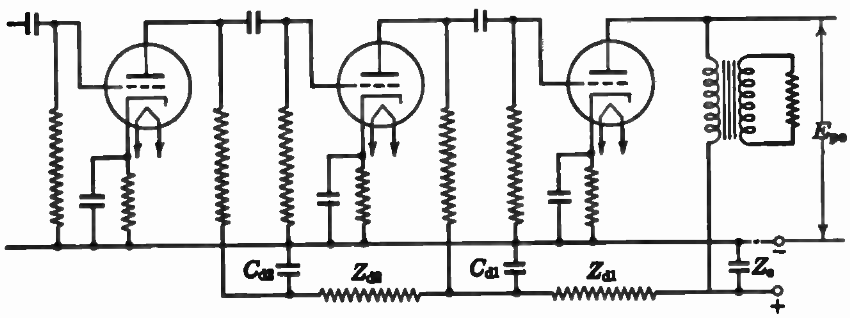


Fig. 3-32. The use of de-coupling circuits in a three-stage amplifier.

circuit. The decoupling factor from the output to the first tube is equal to the product of the decoupling factors of each section of the circuit, i.e. the overall decoupling $D = \frac{R_1R_2}{\omega^2C_1C_2}$. Thus the two stage decoupling circuit provides much more overall decoupling than if the same amount of resistance and capacitance were used in a single decoupling section. In addition the two-section circuit provides decoupling between all stages of the amplifier.

The decoupling circuit also provides additional filtering when a filtered rectifier is used for the plate supply. When the reactance of a decoupling condenser is one-tenth, or less, than a decoupling resistance, the hum, or noise, voltage in the d-c supply will be attenuated 20 Db or more for each section of the decoupling circuit. Hence a two section decoupling circuit as shown in Fig. 3-32 can easily result in 40 Db or more attenuation in the hum voltage from the d-c supply to the input tube of the amplifier. Consequently good filtering is obtained where it is needed, and since the decoupling circuit for each tube is like the type of circuit used for low frequency compensation in a resistance-capacitance coupled amplifier it can also be used for this purpose provided it is properly designed to give compensation as well as decoupling.

Some decoupling can be obtained by connecting the plate circuits of the amplifier to different points on the filter of the d-c supply when a filtered rectifier is used. The input stages should be connected to the better filtered portion of the rectifier.

3-13. Feed-Back from Self-Bias Circuits.- The grid bias for each stage of a Class A amplifier is usually obtained from a resistor R_k in the cathode circuit of the tube. If no feed-back is wanted this resistor is shunted by a condenser with a reactance small compared to R_k for the lowest frequency to be amplified. When some feed-back is wanted only a portion of the resistor R_k is shunted with a capacitance. In either case it is necessary to know how much capacitance is necessary. If the capacitance is too small there will be degenerative feed-back at the lowest frequencies and no feed-back at the higher frequencies.

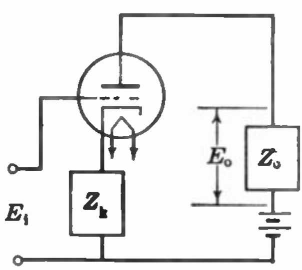


FIG. 3-33. Amplifier with self-bias. Self-bias impedance Z_k is generally a resistor and condenser in parallel.

Referring to Fig. 3-33 let the impedance and voltage symbols be defined as shown. Z_k is the impedance of the parallel circuit of R_k and C_k, or any other impedance that may be placed in the cathode. The voltage from grid to cathode is $\underline{E}_g' = \underline{E}_i + \underline{Z}_k \underline{I}_p$. The total impedance from plate to cathode is $\underline{Z}_o + \underline{Z}_k$ and the plate current is $\underline{I}_p = \dfrac{\mu \underline{E}_g}{r_p + \underline{Z}_k + \underline{Z}_o}$. The voltage across $\underline{Z}_o$ is $\underline{E}_o = -\underline{I}_p \underline{Z}_o$; using these relations and writing $\underline{Y}_k$ for $1/\underline{Z}_k$ there results

$$\frac{\underline{E}_o}{\underline{E}_i} = \frac{(-\mu \underline{Z}_o)}{(r_p + \underline{Z}_o)} \; \frac{1}{1 + \dfrac{1 + \mu}{\underline{Y}_k (r_p + \underline{Z}_o)}}$$

or

$$\frac{\underline{E}_o}{\underline{E}_i} = \begin{bmatrix}\text{Gain with} \\ \text{Fixed Bias} \\ Z_k = 0\end{bmatrix} \left(\frac{1}{1 + \dfrac{1 + \mu}{\underline{Y}_k (r_p + \underline{Z}_o)}} \right) \tag{3-55}$$

In a form more suitable for pentode amplifiers and especially video-frequency amplifiers

$$\frac{\underline{E}_o}{\underline{E}_i} = \begin{bmatrix}\text{Gain with} \\ \text{Fixed Bias} \\ Z_k = 0\end{bmatrix} \left(\frac{1}{1 + \dfrac{g_p + g_{gp}}{\underline{Y}_k (1 + g_p \underline{Z}_o)}} \right) \tag{3-56}$$

For pentodes $g_{gp} \gg g_p$ and $g_p Z_o$ is usually small compared to one, then

$$\frac{\underline{E}_o}{\underline{E}_i} = \begin{bmatrix}\text{Gain with} \\ \text{Fixed Bias} \\ Z_k = 0\end{bmatrix} \left(\frac{1}{1 + \dfrac{g_{gp}}{\underline{Y}_k}} \right) \tag{3-57}$$

From these it is seen that in general to have the gain of the amplifier unaffected by the self-bias impedance $\underline{Z}_k$ the admittance of this circuit must be such that the magnitude of $\underline{Y}_k (1 + g_p \underline{Z}_o) \gg (g_p + g_{gp})$. For pentode amplifiers g_{gp}/Y_k must be small compared to one or $Y_k \gg g_{gp}$. In video-frequency amplifiers the requirements for proper phase shift characteristics, as affected by the self-bias circuit, imposes a more severe limitation on $\underline{Y}_k$ than would be necessary if amplitude distortion alone were concerned.

PART 2. AUDIO-FREQUENCY POWER AMPLIFIERS

3-14. The Single-Tube Class A Power Amplifier. Triode Only.- The Class A power amplifier is operated with such d-c plate and grid voltages that the alternating plate current of each tube has nearly the same wave form as that of the alternating voltage impressed between the grid and cathode. This means that the dynamic I_b vs E_c characteristics of the amplifier must be nearly a straight line over the complete range of the varying grid voltage. In the final analysis, it means the characteristics of the tube itself must be nearly straight lines. The characteristic curves of a tube are straight lines over only very limited ranges of the plate and grid potentials. Hence there is always some distortion. The distortion power in per cent of the total power output increases with the a-c grid voltage, decreases with increase in the output resistance and depends upon where the tube is operated with respect to its plate current versus grid potential characteristics and its plate current versus plate potential characteristics. It is always desirable to operate a power amplifier tube with the least possible d-c plate current and plate potential for a given power output, compatible with the allowable amount of distortion. The d-c grid potential is such that the grid does not become positive on the highest input voltage for which the amplifier is designed.

The characteristic curves of the tube itself can depart quite a little from straight lines in a well-adjusted amplifier before there is a serious departure of the dynamic I_b vs E_c characteristic from a straight line for the higher values of

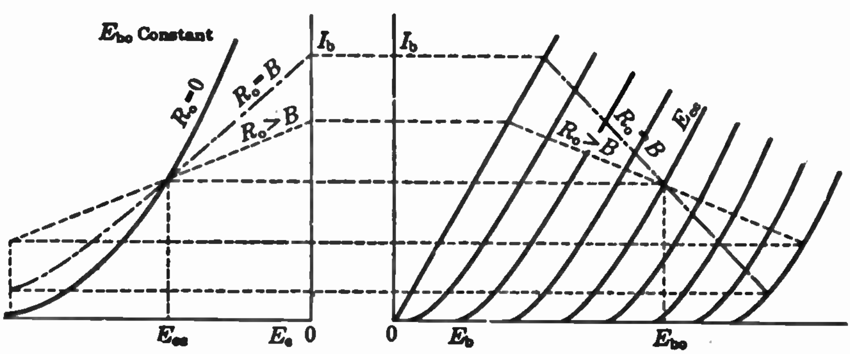


Fig. 3-34. Dynamic I_b vs E_c characteristic of a triode derived from load line. Increasing R_o straightens dynamic characteristic.

load resistance. The reason is that when the grid potential changes the plate potential changes in such a way that μ and r_p are nearly constant. In Fig. 3-34 is shown in a general way how increasing the load resistance straightens out the dynamic characteristic of the amplifier. As the dynamic characteristic becomes straighter the distortion becomes less. Hence increasing R_o decreases the ratio of the distortion power to the total power output for a given input voltage. The power output for a fixed input voltage increases at first with increase in R_o, reaches a maximum when $R_o = r_p$ and then decreases. This reasoning indicates that the highest ratio of total power to distortion power results when R_o is greater than r_p when the input voltage is held constant. Most of the distortion in an amplifier results from driving the grid into the region of very small plate currents where the I_b vs E_c characteristics are far from straight lines or where r_p varies quite rapidly with the operating voltages. Hence in a power amplifier the minimum value of the plate current, i.e. the plate current on the negative crest of the input voltage should not be allowed to fall below a certain specified value. Consequently this limits the alternating grid potential for a fixed set of d-c operating voltages and a given output resistance. It is the purpose of the next section to arrive at the relation between R_o in terms of r_p and the d-c operating voltages for the maximum output power possible for a fixed amount of distortion power in per cent of the total output power.

The Best Values of R_o and E_{cc} for Maximum Power; Triodes Only. It is seen that upon increasing R_o, while E_{cc} and E_{bb} are held constant, the dynamic characteristic becomes straighter. Then for the same amount of distortion E_g can be increased, provided E_{cc} is also adjusted, as R_o is increased. The power output is approximately given by the relation (divide E_o^2 as given by eq. (3-2) by R_o):

$$P_o = \frac{E_g^2 \mu^2 R_o}{(R_o + r_p)^2} \tag{3-58}$$

When E_g is fixed the power output is a maximum when $R_o = r_p$. But since as R_o is increased E_g can also be increased, it is obvious that P_o will be a maximum for some value of R_o greater than r_p. Hence, for a given plate battery potential E_{bb}, the highest power output will be obtained when the peak value of E_g is equal to E_{cc} and when the voltage E_{cc} and the output resistance R_o are such that the plate current on the negative crest of the voltage E_g is just equal to the allowable minimum.

In order to derive the relation for the bias voltage E_{cc} and output resistance R_o which will result in maximum power output when the peak value of E_g is equal to E_{cc}, it is assumed that both the static and dynamic curves are straight lines for

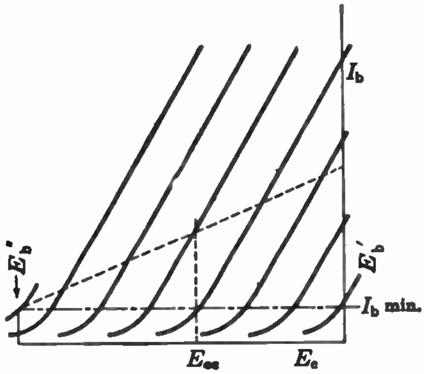


FIG. 3-35. Ideal I_b vs E_c characteristics of a triode. μ and r_p are constant for all values of plate current greater than I_b min.

all values of plate current greater than a certain minimum value. Figure 3-35 illustrates these conditions. In this figure the operating point and load resistance are so adjusted that on the positive crest of the input voltage the grid potential will not go positive and on the negative crest the plate current will not fall below the value I_b min. Hence, for maximum output the peak value of alternating grid voltage should be just equal to the negative grid bias E_{cc}.

In the fundamental theory of the three element tube it was found that a given change in plate current is produced by a change in plate potential which is $-\mu$ times the change in grid potential necessary to produce the same change in plate current. Now referring to Fig. 3-35 the plate potential is E_b'', as shown, when the total grid potential is $2E_{cc}$ and the plate current is I_b min. When the total grid potential is zero, the plate potential is E_b' for the plate current I_b min.

$$I_b = f(E_b + \mu E_c)$$

$$I_b \text{ min.} = f(E_b' + \mu 0)$$

also $$I_b \text{ min.} = f(E_b'' + 2\mu E_{cc})$$

hence $$E_b'' + 2\mu E_{cc} = E_b'$$

Hence $$E_b'' = -2\mu E_{cc} + E_b'$$

or $$E_b'' - E_b' = -2\mu E_{cc}$$

But in the plate circuit of the amplifier, if E_{bo} is the plate d-c potential at the operating point, i.e. $E_{bo} = E_{bb} - I_{bo}R_o$.

$$E_{bo} - R_o i_p = e_b \text{ and } i_p = \frac{\mu e_g}{r_p + R_o}$$

Hence for the maximum value of e_g, i.e., E_{cc}:

$$E_{bo} - \frac{\mu E_{cc} R_o}{r_p + R_o} = E_b''$$

Therefore $$E_{bo} - \frac{\mu E_{cc} R_o}{r_p + R_o} - E_b' = E_b'' - E_b' = -2\mu E_{cc}$$

or finally $$E_{bo} - \frac{E_{cc} R_o}{r_p + R_o} - E_b' = -2\mu E_{cc} \qquad (3\text{-}59)$$

The above equation furnishes all the information necessary for adjusting the voltages E_{cc} and E_g and the load resistance R_o so that maximum power output will be obtained.

The solution of eq. (3-59) for the grid bias voltage E_{cc} gives

$$E_{cc} = -\frac{E_{bo} - E_b'}{\mu}\left(\frac{R_o + r_p}{R_o + 2r_p}\right) \tag{3-60}$$

where E_b' is the plate voltage required to produce the minimum plate current when the total grid potential is zero.

The solution of eq. 3-59 for the resistance R_o gives

$$R_o = -r_p\left[1 + \frac{E_{cc}}{\frac{E_{bo} - E_b'}{\mu} + E_{cc}}\right] \tag{3-61}$$

which is the output resistance to be used for any grid bias E_{cc}.

The rms value of the voltage across the output resistance is

$$E_o = \frac{1}{\sqrt{2}}\,\frac{E_{cc}R_o}{r_p + R_o}$$

when the peak value of $E_g = E_{cc}$.
This voltage squared and divided by R_o gives the power output. Hence

$$\text{Power output} = P_o = \frac{\mu^2 E_{cc}{}^2 R_o}{2(r_p + R_o)^2} \tag{3-62}$$

Equation (3-62) may also be written in terms of the voltages E_{bo} and E_b'. If the best possible value of grid bias is used, i.e., the value given by eq. 3-60, the power output becomes

$$P_o = \frac{(E_{bo} - E_b')^2}{2}\,\frac{R_o}{(R_o + 2r_p)^2} \tag{3-63}$$

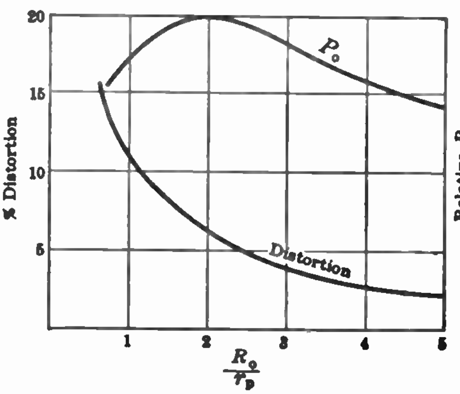


FIG. 3-36. Typical dependance of power output and distortion on R_o/r_p for a triode.

Under these conditions it is to be remembered that the peak value of the input voltage must be equal to the grid bias voltage E_{cc}. In order to find the load resistance R_o which will result in maximum power output when R_o is adjusted, eq. 3-63 is differentiated with respect to R_o, set equal to zero and solved for R_o. This procedure results in the fact that R_o must equal $2r_p$. Hence the maximum power output without distortion is

$$P_o \max = (E_{bo} - E_b')^2 \frac{1}{16r_p} \tag{3-64}$$

without distortion:

when $R_o = 2r_p$

$$E_{cc} = 3/4 \frac{(E_{bo} - E_b')}{\mu}, \quad \text{then } P_o = \frac{2}{9} \frac{\mu^2 E_g^2}{r_p}$$

where $\sqrt{2}E_g = E_{cc}$

E_{bo} is the plate voltage at the operating point.
E_{bb} = Battery voltage = R_o times (Plate current at operating point)
E_b' is the voltage required on the plate to produce the minimum plate current allowable when the total grid potential is zero.
E_g is the rms value of the a-c input voltage. It is to be remembered that the peak of the alternating voltage input must be equal to the grid bias voltage E_{cc} and that the grid is not to be driven positive. For amplifiers of this class the grid potential is not allowed to become positive because of grid current.

Departure from Ideal Conditions. In deriving the above relations ideal conditions were assumed, i.e., above a certain minimum plate current the static characteristics were straight lines. Now actually tubes do not have such ideal characteristics. Over the entire range of voltages there is always some curvature in the characteristics which will give rise to harmonic distortion. This curvature increases as the plate current decreases. Hence, in specifying a minimum value for the plate current the permissible amount of total harmonic content must be considered. Consequently, the theory for the ideal case is somewhat modified. Particularly, the best value of output resistance is very seldom equal to $2r_p$ but is usually somewhat greater. This follows from the statement made earlier that increasing R_o straightens out the dynamic characteristic and thereby reduces distortion.

Matching Load Impedance to Tube. Usually any utilization device into which the power is delivered is coupled into the plate circuit by means of a transformer. When the utilization device is a resistance, the ratio of turns of the transformer should be such that the load resistance referred to the primary side is equal to the resistance specified for the tube. The losses and leakage inductances of the transformer are kept within certain limits, because they play very important parts in the overall performance of the combined tube and transformer.

When the d-c resistance of the primary winding of the transformer is only a few per cent of the transformed load resistance, the potentials E_{bo} and E_{bb} are substantially the same. The load line on the ideal I_b versus E_b chart is substantially the same as the load line which would be obtained by placing a resistance R_o equal to the transformed load resistance directly in the plate circuit with a plate source potential equal to the d-c potential of the plate plus the d-c potential drop in the resistance R_o. On an actual I_b versus E_b chart the load line does not in general pass through the point I_{bo}, E_{bo} when the output resistance has a different value to d-c and to a-c. Hence in general if the load line is drawn through point I_{bo}, E_{bo} when the a-c and d-c load resistances are not equal there will be some error in arriving at the power output and distortion from graphical analysis.

3-15. The Single-Tube Class A Power Amplifier. Pentodes and Beam Tubes.- The load resistance requirements for Class A pentode and beam-tube power amplifiers are quite different from those of a Class A triode-power amplifier. Pentode and beam tubes have load resistance requirements somewhat similar to each other. The material difference between the I_b vs E_b characteristics of a pentode and of a beam tube is that the transition from a plate current which changes rapidly with E_b to one which changes slowly with E_b occurs sharper and at a lower plate potential in a beam tube than in a pentode. In either

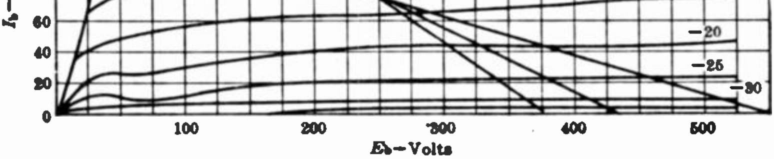


FIG. 3-37. Load lines on the $I_b - E_b$ chart for a beam power amplifier.

tube the load resistance requirement for a given distortion is much lower than the plate resistance of the tube. Figure 3-38 illustrates the effect of load resistance on the dynamic I_b vs E_c characteristics of a typical Class A beam-tube power amplifier. For this tube a 2500-ohm load resistance gives the minimum per cent distortion when operating with 250 V. on the plate,

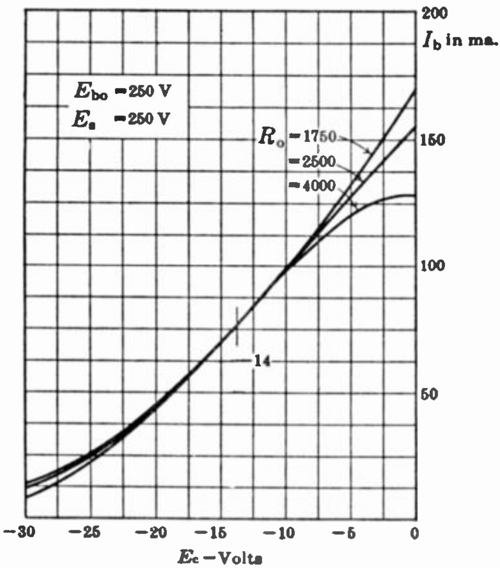


FIG. 3-38. The dynamic $I_b - E_c$ characteristics of the beam power amplifier of Fig. 3-37.

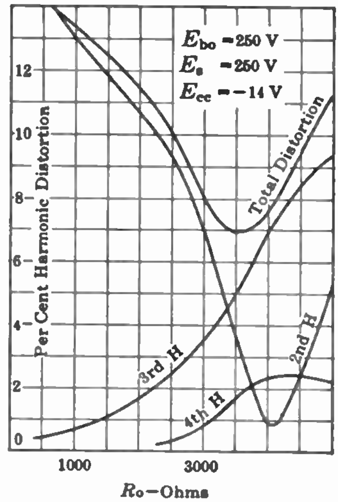


FIG. 3-39. Per cent distortion vs load resistance for the beam power amplifier of Fig. 3-37.

250 V. on the screen, -14 V. on the grid, and 10 rms volts a-c input. Under these same conditions of operation the second harmonic distortion is a minimum when R_o = 4000 ohms. A 2500-ohm load resistor results in 10 per cent total distortion with a power output of 6.5 watts. For load resistances greater than 3500 ohms the total per cent distortion rises so rapidly that there is actually a reduction in fundamental power output, although the load resistance is still well below the plate resistance of the tube. When this tube is operated in a Class A balanced push-pull amplifier for a distortion of two per cent or less, the load resistance to each tube should be about 1500 ohms. This would result mostly in third harmonic distortion because, as will be seen later, the even harmonics will balance out across the entire load when the load resistors are placed directly in the plate circuit or in single non-inductive load when the load is coupled to the tube by means of a transformer.

Single-ended beam and pentode power amplifiers have slightly higher distortion than a triode power amplifier. In the pentode, the third-harmonic distortion is particularly bad. Because minimum distortion occurs for a load resistance which is small compared to the plate resistance in a pentode or beam-power amplifier, whereas in a triode the distortion decreases as the load resistance is increased, the triode tube causes less distortion in those cases where the load resistance increases with the frequency of operation. The efficiency of power conversion is slightly higher for pentode and beam-power amplifiers than it is for a triode amplifier. Also pentode and beam-power amplifiers require less driving voltage than a triode power amplifier because of the higher amplification factor of these tubes which in turn is due to the action of the screen which is maintained at a constant d-c potential nearly equal to the d-c potential of the plate.

3-16. Push-Pull Audio-Frequency Power Amplifiers.- When two tubes are arranged in a power- or voltage-amplifier circuit in such a manner that the input voltage to one tube is 180° out of phase with the input voltage to the other tube, and the plate circuits are so arranged that the output voltages from each plate to cathode are 180° out of phase the arrangement is called a push-pull amplifier. Figure 3-40 shows how the tubes are arranged in such an amplifier. The secondary of the input transformer and the primary of the output transformers are each equivalent to continuous windings with a center tap connection as shown. The connections are such that when the voltage to the grid of tube A is rising above the d-c bias the voltage to the grid of B is falling below the d-c bias. Under ideal conditions the result is that the fundamental or any odd harmonic of an alternating component of current of tube B adds to the corresponding component of tube A in the primary of the output transformer. This produces an alternating current in the output resistance R_L.

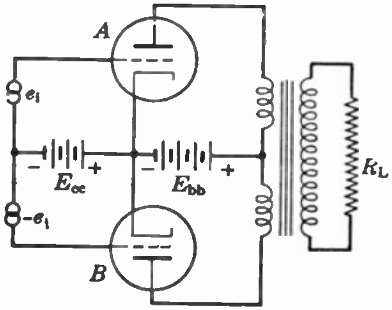


Fig. 3-40. Schematic circuit of a push-pull amplifier.

The push-pull arrangement has certain advantages over connecting the two tubes in a straight parallel arrangement. One obvious advantage is that when the two tubes are exactly alike there is no d-c flux in the core of the output transformer; d-c flux in the core of a transformer lowers the effective a-c permeability of the core material. The greater advantage is the reduction in the distortion components in the a-c current in the output resistance R_0. This advantage will be shown in more detail in the discussion of this type of amplifier.

Push-pull audio-frequency power amplifiers are classified as to their mode of operation into Class A, Class AB, and

Class B, according to the definitions given in Sec. 3-1 of this chapter. In reality there is no sharp line of demarcation between these three classifications. A theory will be developed which will comprise all three. This theory, partly analytical and partly graphical, will be derived from the combined characteristics of the two tubes.

The Composite Characteristics of the Push-Pull Amplifier. The two tubes in push-pull may be replaced by a single equivalent tube which has for its I_b versus E_b characteristics the algebraic sum of the characteristics of the two tubes. This method of approach gives all the necessary information in so far as a-c voltages and currents are concerned. The graphical method of approach to the idea of a single equivalent tube is as follows: For two identical tubes two sets of I_b versus E_b curves are drawn one above the other which is inverted so that voltage increases to the right on the upper set and to the left on the lower set. These two sets are so adjusted with respect to each other that they coincide at the d-c operating plate voltage. Then an I_b versus E_b curve for a grid voltage $E_{cd} + \Delta E_c$ of tube A is added to a similar curve for a grid voltage of $E_{cd} - \Delta E_c$ of tube B. This is done for as many pairs of curves as are necessary to carry out the graphical analysis. The resultant characteristics are called composite characteristics. Figure 3-41 illustrates how these composite characteristics are obtained. The dashed lines are the composite curves.

The set of composite characteristics represents the I_b versus E_b characteristics which a single equivalent tube must have in order to replace the two tubes in push-pull. All the information, except the d-c plate current, which is pertinent to the analysis of the operation of the two tubes in push-pull may be obtained from the set of composite curves. That these curves can furnish this information may be justified as follows: Assume first that the grid a-c voltage is zero. When the instantaneous plate voltage to the upper tube is above the d-c operating voltage by an amount ΔE_b volts the instantaneous plate voltage of the lower tube is ΔE_b volts below the d-c value. The instantaneous plate current of the upper tube is found on the corresponding I_b versus E_b curve for that tube and is the current I_{bA}. The plate current for the lower tube is I_{bB}. Now at normal plate voltage the two equal plate currents are flowing in opposite directions in the primary of the output transformer. Hence, there is no magnetic field in the core. With the new voltage $E_{bd} + \Delta E_b$ on the upper tube and $E_{bd} - \Delta E_b$ on the lower tube the net current in the primary of the output transformer is the algebraic sum of the two plate currents at the new plate voltages. This is the current I_{bc} taken from the composite characteristic at the plate voltages $E_{bd} + \Delta E_b$ for the upper tube and $E_{bd} - \Delta E_b$ for the lower tube. This is the current

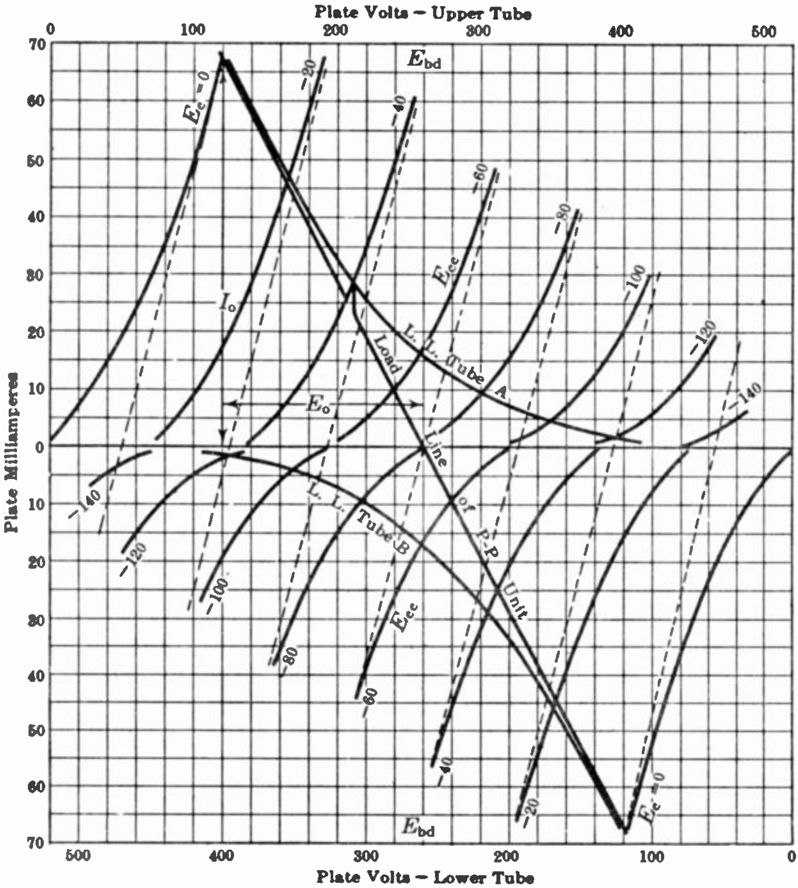


Fig. 3-41. The composite characteristics of a push-pull amplifier.

which must be multiplied by the turns of one winding of the primary to give the ampere turns which produces the instantaneous flux in the core. If N_p and N_s are the number of turns in the primary and secondary windings respectively the load current $I_L = I_{bc}N_p/N_s$. This is strictly true only for an ideal transformer and practically true for any well-designed transformer.

Graphical Solution of a Push-pull Amplifier. The method for establishing the I_b versus E_b characteristics of the single tube which is the equivalent to the two tubes in push-pull was given in the preceding section. The method for arriving at the form of the plate current versus grid potential will now be given. In Fig. 3-41 is shown the I_b versus E_b characteristics of the two tubes arranged also to show the composite characteristics. If the load resistor is R_L ohms and the ratio of the primary turns of one winding to the secondary turns of the output transformer is N, the output resistance for the equivalent single tube is N^2R_L ohms. Hence, across the composite I_b vs E_b characteristics is drawn the load line with a negative slope of $1/N^2R_L$. This is the load line of the push-pull combination. It must be drawn through the proper d-c operating point. The points where this load line crosses the composite characteristics are the values of the current in the transferred load resistance which are due to the particular grid voltages for which the composite characteristics are drawn.

The load line of each tube is a line drawn through the points of intersection of the vertical lines, which are in turn drawn through the intersections of the composite characteristics and composite load line, and the I_b versus E_b characteristics of the tube. These are the curved lines shown in Fig. 3-41 and are obtained as illustrated. When the I_b versus E_b characteristics are straight lines over the complete range of operation the load lines of each tube are straight but have only 1/2 the negative slope of the composite load line. In general the load lines of the individual tubes are not straight lines because the instantaneous plate voltage of either tube is not due entirely to the action of that tube alone but is due to the combined action of the two tubes. There is no even harmonic distortion in the output voltage or the voltage across either half of the primary of the transformer. Consequently the load line of either tube must be of such a nature that the dynamic relation between E_b and E_c does not contain even harmonic distortion. In another way, the situation is such that the plate current of either tube contains both even and odd harmonics whereas the total plate voltage contains only odd harmonics. This situation demands a load line for either tube that is not a straight line when operation takes place over the non-linear portions of the I_b vs E_b characteristics.

The graphical construction shown in Fig. 3-41 gives a correct picture only when the two tubes are exactly alike and when the transformer is ideal and the d-c resistance offered to each tube is negligible compared to the a-c resistance. When the d-c resistance of the transformer is not negligibly small the d-c potential of each plate, for any given a-c input voltage differs from the source potential E_{bb}. Then the I_b vs E_b characteristics of the two tubes have to be joined at the d-c plate operating potential E_{bd}. Due to partial plate current rectification the d-c plate current I_{bd} is not equal to the quiescent d-c plate current I_{bo}. Consequently it is difficult to arrive at the true dynamic d-c plate potential E_{bd} and the true graphical picture of operation. The graphical construction for the Class B amplifier is exceptionally bad because of the large amount of plate current rectification. Self-bias also produces errors in the graphical picture because the grid-bias of either tube, i.e. the voltage E_{cd} depends upon the total dynamic d-c plate current I_{bd}. Self-bias is not generally used when there is a large difference between the dynamic and quiescent d-c plate currents as in Class B operation.

The Fundamental Power Output and Efficiency of a Push-pull Amplifier. The fundamental power output of the push-pull amplifier is obtained by considering the push-pull arrangement to be replaced by a single tube that has I_b vs E_b characteristics exactly like the composite characteristics of the two push-pull tubes. This equivalent tube will have a plate conductance equal to the slope of the composite characteristic. The amplification factor of the equivalent tube will be practically the same as the amplification factor of either one of the push-pull tubes. If the plate conductance is symbolized by g_p' the plate resistance will be $r_p' = 1/g_p'$. Now, as stated before, the output resistance is N^2R_L where N is the ratio of the turns of one primary winding to the secondary turns. The grid voltage of the equivalent tube will be equal to the grid voltage of only one of the push-pull tubes. Consequently the power output of the push-pull amplifier is

$$P_o = \frac{\mu^2 E_g^2}{(N^2R_L + r_p')^2} N^2R_L \qquad (3\text{-}65)$$

When the push-pull tubes are operated as Class A amplifiers r_p' is practically equal to $r_p/2$. When the tubes are operated as Class B amplifiers then r_p' is practically equal to r_p. For Class AB operation r_p' lies somewhere between $r_p/2$ and r_p.

The fundamental power output can also be obtained from the graphical harmonic analysis. It is equal to the fundamental component of the current for the push-pull unit squared times the output resistance N^2R_L.

The order of plate circuit efficiencies of the three classes of push-pull amplifiers are as follows: Class B most efficient, Class AB next, and Class A least. The reason is that in the Class B amplifier d-c current, except for a very small amount, flows only when the a-c input voltage is impressed in the grid circuits and is proportional to the input voltage. It is a little greater than 0.636 times the peak value of the a-c output current of the unit. In Class A operation the d-c plate current is equal to or greater than the peak value of the a-c output current of the unit. Consequently if two tubes are operated first as Class A, then as Class B, at the same d-c plate voltages and adjusted to give the same power output, the Class B amplifier will require less d-c power input to the plate. Typical efficiencies for the three classes, when operated so that the harmonic distortion is not excessive, are Class A 30 per cent, Class AB 40 per cent, and Class B 70 per cent.

The Distortion Components in a Push-pull Amplifier. The dynamic I_b versus E_c characteristic of each tube is derived in the usual manner from the I_b vs E_b characteristics and the load line for the tube. The composite dynamic characteristic may either be derived as the algebraic sum of the individual dynamic characteristics as shown in Fig. 3-42 or it may be derived directly from I_b vs E_b composite characteristics and the composite

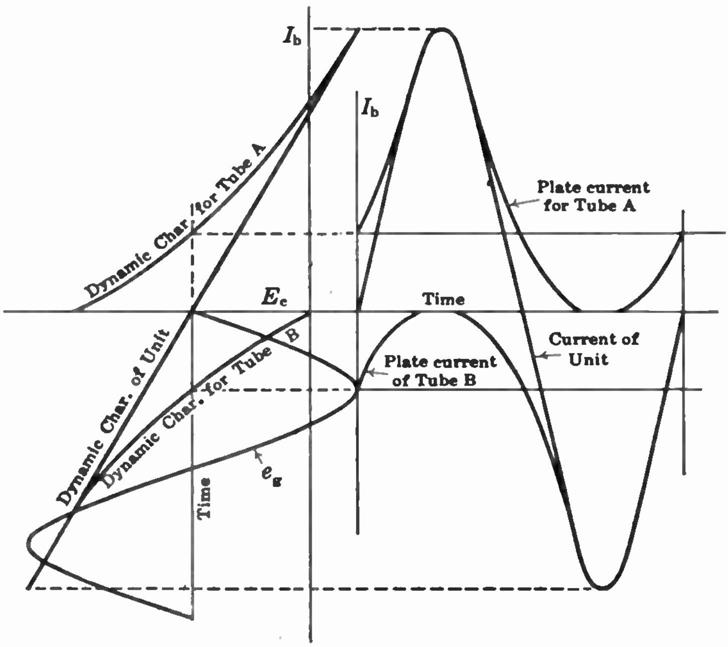


FIG. 3-42. The $I_b - E_c$ dynamic characteristics of a push-pull amplifier.

load line of the push-pull unit.

It is evident that if the two tubes are exactly alike the instantaneous output current curve of the unit which is derived from the composite dynamic characteristic will be symmetrical about the zero line. Hence, under these conditions there will be no even harmonic distortion in the load resistance. There will be only odd harmonic distortion. In practice the two tubes will not be exactly alike and also the transformer will not be ideal. Hence, there will, in general, be some even as well as odd harmonic distortion in the load resistance. The graphical method for evaluating the harmonic distortion which is given in Sec. 3-17 can be applied to the push-pull case by evaluating the various currents involved from the dynamic I_b vs E_c characteristics of the unit. In order to get the total d-c plate current the graphical harmonic analysis has to be applied to the dynamic I_b vs E_c characteristic of the individual tubes.

Power Input to Class AB_2 and Class B Amplifiers. When the grid of an amplifier tube is driven positive with respect to the cathode grid current flows for the fraction of the input cycle that the grid is positive. Each resulting grid-current pulse that occurs for each positive half cycle of the input voltage is nearly like the top portion of a sine-squared function. The grid current is made up of a fundamental and several harmonics. The fundamental component I_{g1} times the input voltage, assumed to be a sine function, gives the power required to drive the tube. The fundamental component of grid current can be evaluated from graphical analysis exactly as is done in the case of a Class C or Class B radio-frequency amplifier. This analysis is given in Sec. 3-20.

Because grid current flows for only a fraction of the positive half cycle the source driving the grid should have good voltage regulation. That is, the driving source should be equivalent to a low impedance compared to the input impedance of the grid when current is flowing. If this is not the case the driving voltage will be distorted. To achieve the equivalence of a fairly low impedance source generally requires transformer coupling in the driver stage and a driver tube capable of furnishing more power than is actually required to drive the Class B amplifier. The transformer reduces the driver tube impedance to its secondary side.

3-17. Graphical Method for Evaluating the Harmonics in a Power Amplifier When the Load is a Pure Resistance.- Usually, the presence in the output of an amplifier of harmonics of a sinusoidal input voltage represents distortion. Harmonic distortion D is defined as

$$D = \frac{(I_2^2 + I_3^2 + \dots + I_n^2)^{1/2}}{I_1}$$

where I_1 is the amplitude of the fundamental and $I_2, I_3, \dots,$ I_n the amplitudes of the second, third, etc. harmonics (I.R.E. Standards on Electronics, 1938, p. 45). It is thus desirable to have available methods for determining harmonic content. One such method is outlined in the following. Experimental methods are discussed in the I.R.E. Standards.

When the load impedance of a power amplifier is a pure resistance the instantaneous plate current is such that there is symmetry of current on the positive half of the input cycle about the current which flows when the input voltage is a maximum, and symmetry on the negative half of the input cycle about the current which flows when the input voltage is a minimum. This is because the path of operation on the I_b vs E_c characteristics is a single valued function and therefore i_b amperes produced at an angle θ' less than $\pi/2$ from zero time on the input cycle is reproduced when the angle is $\pi - \theta'$. These properties are illustrated by Fig. 3-43.

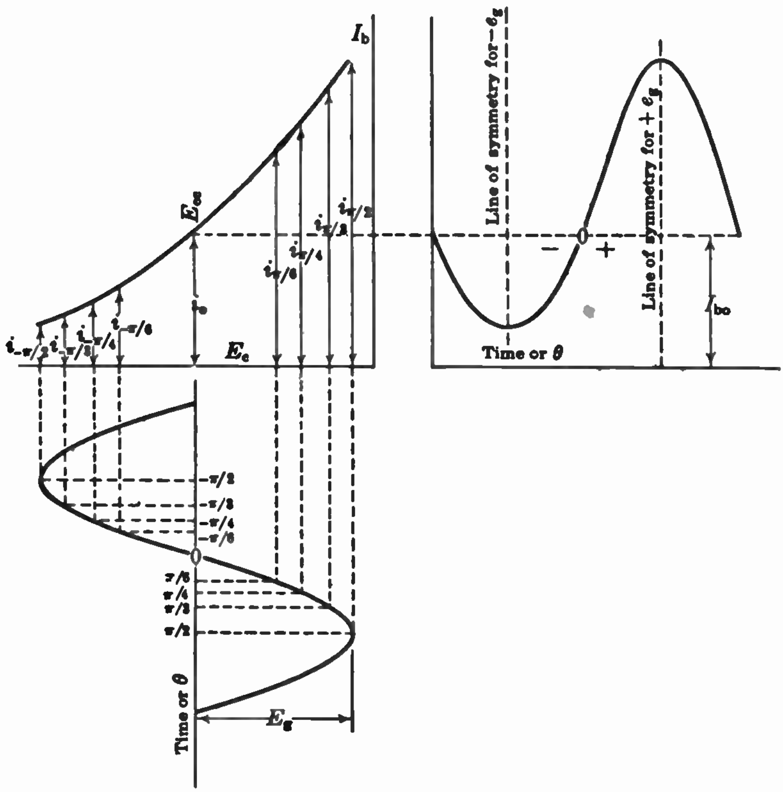


FIG. 3-43. Graph showing the method for scaling the necessary currents from the dynamic $I_b - E_c$ curve for evaluating the harmonics in i_p.

When a pulsating current has the properties described and illustrated above it can be represented by a constant term plus sine terms of odd harmonics plus cosine terms of even harmonics.[8] Hence under these conditions of operation the total plate current is given by

8. See the section on Fourier Series in Ch. 1.

$$i_b = I_{bd} + I_1 \sin \omega t + I_2 \cos 2\omega t + I_3 \sin 3\omega t + I_4 \cos 4\omega t + I_5 \sin 5\omega t + I_6 \cos 6\omega t + I_7 \sin 7\omega t + \ldots \quad (3\text{-}66)$$

The coefficients I_{bd}, I_1, I_2, etc. are the peak amplitudes of the constant term, the fundamental frequency term, the second harmonic frequency term, etc. respectively. These coefficients, or peak amplitudes, can be evaluated by properly selecting points on the input emf wave and writing the total current i_b for each of these points. The minimum number of such points is equal to the number of coefficients which is desired in any given analysis. This procedure gives the following expressions for the d-c current I_{bd} and the first seven harmonic currents.

$$I_0 = I_{bd} + I_2 + I_4 + I_6 - I_{bo}.$$

$$i_{\pi/6} = I_{bd} + 0.5I_1 + 0.5I_2 + I_3 - 0.5I_4 + 0.5I_5 - I_6 - 0.5I_7.$$

$$i_{-\pi/6} = I_{bd} - 0.5I_1 + 0.5I_2 - I_3 - 0.5I_4 - 0.5I_5 - I_6 + 0.5I_7.$$

$$i_{\pi/4} = I_{bd} + 0.707I_1 + 0.707I_3 - I_4 - 0.707I_5 - 0.707I_7.$$

$$i_{-\pi/4} = I_{bd} - 0.707I_1 - 0.707I_3 - I_4 + 0.707I_5 + 0.707I_7.$$

$$i_{\pi/3} = I_{bd} + 0.866I_1 - 0.5I_2 - 0.5I_4 - 0.866I_5 + I_6 + 0.866I_7.$$

$$i_{-\pi/3} = I_{bd} - 0.866I_1 - 0.5I_2 - 0.5I_4 + 0.866I_5 + I_6 - 0.866I_7.$$

$$i_{\pi/2} = I_{bd} - I_1 - I_2 - I_3 + I_4 + I_5 - I_6 - I_7.$$

$$i_{-\pi/2} = I_{bd} - I_1 - I_2 + I_3 + I_4 - I_5 - I_6 + I_7.$$

where the fractional π subscripts and 0 refer to angles of θ on the a-c input emf measured from zero reference time which occurs when the input voltage starts to raise the grid potential above the bias potential E_{co}. With this notation the grid potentials measured from 0 which cause each of the above currents on the left sides of the above expressions are as follows. (It must be remembered that E_{co} may have a positive or negative numerical value, generally negative for all cases except in some cases of Class B operation, when $E_{co} = 0$. E_g is the peak value of the a-c input voltage.)

Plate Current		Grid Potential
i_{bo}	------	E_{co}
$i_{\pi/6}$	------	$E_{co} + 0.5E_g$
$i_{-\pi/6}$	------	$E_{co} - 0.5E_g$
$i_{\pi/4}$	------	$E_{co} + 0.707E_g$
$i_{-\pi/4}$	------	$E_{co} - 0.707E_g$
$i_{\pi/3}$	------	$E_{co} + 0.866E_g$

Plate Current		Grid Potential
$i_{-\pi/3}$	------	$E_{co} - 0.866E_g$
$i_{\pi/2}$	------	$E_{co} + E_g$
$i_{-\pi/2}$	------	$E_{co} - E_g$

Consequently the currents i_{bo}, $i_{\pi/2}$ etc. are evaluated by locating each of instantaneous grid potentials shown in the table and scaling the currents on the dynamic I_b vs E_c characteristic at each of these points.

The equation for i_{bo}, $i_{\pi/2}$ etc. are solved for each of the currents I_{bd}, I_1, I_2 etc. One method for solving these equations is to first write certain sums and differences. For the sake of brevity the sums and differences are given the following symbols:

$$i_{\pi/6} + i_{-\pi/6} = A, \qquad i_{\pi/6} - i_{-\pi/6} = B, \tag{3-68}$$

$$i_{\pi/4} + i_{-\pi/4} = C, \qquad i_{\pi/4} - i_{-\pi/4} = D,$$

$$i_{\pi/3} + i_{-\pi/3} = E, \qquad i_{\pi/3} - i_{-\pi/3} = F,$$

$$i_{\pi/2} + i_{-\pi/2} = G \quad \text{and} \quad i_{\pi/2} - i_{-\pi/2} = H$$

Using this notation there results,

$$A = 2I_{bd} + I_2 - I_4 - 2I_6 \tag{3-69}$$

$$B = I_1 + 2I_3 + I_5 - I_7$$

$$C = 2I_{bd} - 2I_4$$

$$D = 1.41I_1 + 1.41I_3 - 1.41I_5 - 1.41I_7$$

$$E = 2I_{bd} - I_2 - I_4 + 2I_6$$

$$F = 1.73I_1 - 1.73I_5 + 1.73I_7$$

$$G = 2I_{bd} - 2I_2 + 2I_4 - 2I_6$$

$$H = 2I_1 - 2I_3 + 2I_5 - 2I_7$$

Note that the sum terms are made up of even harmonics and the d-c term while the difference terms are made up of only odd harmonics. From these equations there result

$$4I_{bd} - 2I_4 = A + E$$

and

$$2I_{bd} - 2I_4 = C$$

By subtraction

$$2I_{bd} = A + E - C$$

or

$$I_{bd} = \frac{A + E - C}{2}$$

Also

$$I_4 = \frac{A + E}{2} - C$$

Since $$I_{bo} = I_{bd} + I_2 + I_4 + I_6$$

and from the above solutions for I_4 and I_{bd} there results

$$I_4 + I_{bd} = A + E - \frac{3}{2} C$$

Then $$I_2 + I_6 = I_{bo} - A - E + \frac{3}{2} C$$

From the equations for A and E there results

$$2I_2 - 4I_6 = A - E$$

These last two equations can be solved for I_2 and I_6:

$$I_2 = \frac{2}{3} I_{bo} - \frac{A}{2} - \frac{5}{6} E + C$$

and $$I_6 = \frac{I_{bo}}{3} - \frac{A}{2} + \frac{C}{2} - \frac{E}{6}$$

This completes the solution for the average current I_{bd} and the even harmonics I_2, I_4, and I_6.

Returning again to the A, B, C, D, etc. equations, it is seen that

$$2B - H = 6I_3$$

or $$I_3 = \frac{B}{3} - \frac{H}{6}$$

It is also seen that

$$4I_1 = 1.154F + H + 2I_3$$

Substituting in this equation the above value for I_3 and dividing by 4 yields

$$I_1 = 0.289F + \frac{1}{6} H + \frac{1}{6} B$$

From the equations for H and D there results

$$-4I_3 + 4I_5 = H - 1.41D$$

Substituting in this equation the expression for I_3 and solving for I_5 there results

$$I_5 = 0.083H - 0.352D + 0.333B$$

Finally from the equations for B and H

$$-3I_7 + 3I_1 + 3I_5 = B + H$$

or $$I_7 = I_1 + I_5 - (B/3 + H/3)$$

But $$I_1 + I_5 = 0.289F + 1/4H + 1/2B - 0.352D$$

Hence

$$I_7 = 0.289F - 1/12H + 1/6B - 0.352D$$

Summarizing:

$$I_{bd} = 0.5A + 0.5E - 0.5C = \text{D-C} \quad (3\text{-}70)$$

$$I_1 = 0.289F + 0.166H + 0.166B = \text{Fundamental A-C}$$

$$I_2 = 0.666I_{bo} - 0.5A - 0.833E + C = \text{2nd harmonic}$$

$$I_3 = 0.333B - 0.166H = \text{3rd harmonic}$$

$$I_4 = 0.5A + 0.5E - C = \text{4th harmonic}$$

$$I_5 = 0.083H - 0.352D + 0.333D = \text{5th harmonic}$$

$$I_6 = 0.333I_{bo} - 0.5A + 0.5C - 0.166E = \text{6th harmonic}$$

$$I_7 = 0.289F - 0.083H + 0.166B - 0.352D = \text{7th harmonic}$$

where the a-c components are all peak values. Consequently the procedure for determining the first seven harmonic currents and the d-c current in a load resistance R_o consists of first drawing on suitable graph paper the I_b vs E_c dynamic characteristic which is derived from the load line on the I_b vs E_b characteristics. Then the voltages, in terms of E_g and E_{cc}, which give the plate currents I_o, $i_{\pi/6}$, $i_{-\pi/6}$ etc., are located on the E_c axis according to the table given above. The currents I_o, $i_{\pi/6}$, $i_{-\pi/6}$ etc. are then scaled and their values placed in the equations for the current components. These equations are then solved for the d-c and a-c current components.

This graphical method for evaluating the fundamental and harmonic currents can be used for the push-pull power amplifier with a good degree of accuracy. The method given is based on the load resistance being placed directly in the plate circuit of a single-ended amplifier. When the load resistance is transformed into the plate circuit as it most always is in either single-ended or push-pull amplifiers, there is some discrepancy between the true load line and the apparent load line on the I_b - E_b chart. This is due to the difference between the a-c and d-c resistances in the plate circuit. When the d-c resistance is only a few per cent of the a-c resistance the discrepancy is small and the apparent load line can be used without much error. For the push-pull case the load line of the equivalent Class A tube is used for getting the different currents involved either directly or through the dynamic curve of I_b versus E_c. The d-c current can be obtained only by analyzing the individual dynamic curves of the tubes.

PART 3. RADIO FREQUENCY VOLTAGE AMPLIFIERS

3-18. Radio-Frequency Voltage Amplification.- The radio-frequency voltage amplifier is used for obtaining voltage gain in the radio-frequency spectrum. Power amplification is not of primary importance but of course some power conversion takes place in the plate circuit of each stage of the amplifier. Radio-frequency amplifiers are usually of the selective or tuned type and are used for amplifying a narrow band of frequencies. Figure 3-44 illustrates a typical gain versus frequency characteristic of a tuned radio-frequency voltage amplifier. Except between the frequencies f_1 and f_2 the gain is very low. The ideal characteristic is one that discriminates very sharply against frequencies outside the pass band and has a constant gain over the pass band. Such ideal characteristics are never realized but an approach is found in the doubly-tuned amplifier sometimes known as a band-pass amplifier. The tubes of a radio-frequency voltage amplifier are usually operated as Class A because efficiency is not important.

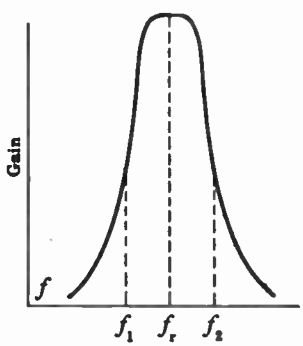


FIG. 3-44. A typical gain vs frequency characteristic of a tuned amplifier.

Pentodes that have their control grids rather completely shielded from the plates are usually used in a radio-frequency voltage amplifier because they yield higher gain and do not require neutralization of the grid-to-plate capacitance. However care must be exercised in the layout of the amplifier or the capacitance and magnetic coupling between the plate and grid of a tube will defeat the advantage of using a pentode. Consequently in radio-frequency amplifiers the tubes and circuits are shielded from each other. The shielding around the coils should not introduce excessive losses and thereby lower their Q factors.

Three simple types of tuned radio-frequency amplifiers are shown in Fig. 3-45. They are shown in the order of their desirable selective characteristics for passing a narrow band of frequencies and eliminating frequencies outside the pass band. It also so happens that the degree of complexity of analysis is in this same order. The type shown in Fig. 3-45a is so inferior to the other two that it is seldom used for a voltage amplifier but is used extensively in power amplifiers.

Single-tuned Circuit Type. The circuit is a parallel circuit of inductance and capacitance placed directly in the plate circuit as shown in Fig. 3-45a. A coupling condenser and grid-leak resistor are used for connecting to the grid of the output tube. Using the symbols shown in the figure the expression for the voltage amplification, when the reactance of the coupling condenser is negligible, is

$$\underline{A} = \frac{-g_{gpi}}{g_{pi} + G_{go} + \frac{R}{R^2 + \omega^2L^2} + j(\omega C - \frac{\omega L}{R^2 + \omega^2 L^2})} \qquad (3\text{-}71)$$

where $\frac{R}{R^2 + \omega^2L^2} + j(\omega C - \frac{\omega L}{R^2 + \omega^2L^2})$ is the admittance of the parallel circuit. When $R \ll \omega L$, $G_{go} \ll \frac{R}{\omega^2 L^2}$, and a pentode is used for the input tube the expression becomes

$$\underline{A} \cong \frac{-g_{gpi}}{\frac{R}{\omega^2L^2} + j(\omega C - \frac{1}{\omega L})} \qquad (3\text{-}72)$$

If the circuit is tuned to the operating frequency by adjusting C the maximum gain is

$$A_{max} \cong \frac{\omega_r^2 L^2}{R} = g_{gpi} Q_r \omega_r L = \frac{L}{RC} \qquad (3\text{-}73)$$

where $\omega_r = \frac{1}{\sqrt{LC}}$ and $Q_r = \frac{\omega_r L}{R} = \frac{1}{R}\sqrt{\frac{L}{C}}$

The gain versus frequency characteristics are shown by the following expression

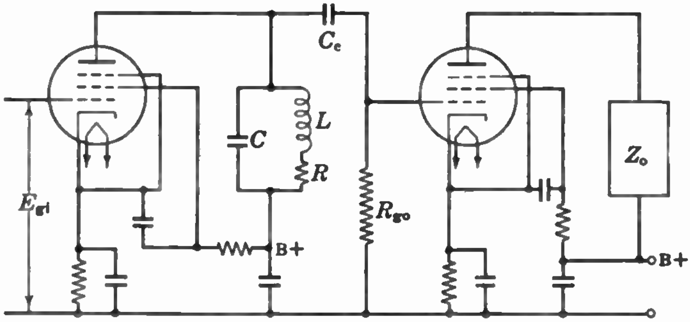
$$\frac{A_{max}}{A \text{ at } f} = \frac{f_r}{f}\sqrt{\frac{f_r^2}{f^2} + Q_r^2\left(\frac{f^2}{f_r^2} - 1\right)^2} \qquad (3\text{-}74)$$

The above expression is somewhat similar to the expression for the ratio of the current at resonance to the current off resonance in a series circuit[9] of L, R, and C, and the two expressions are alike when Q_r is large.

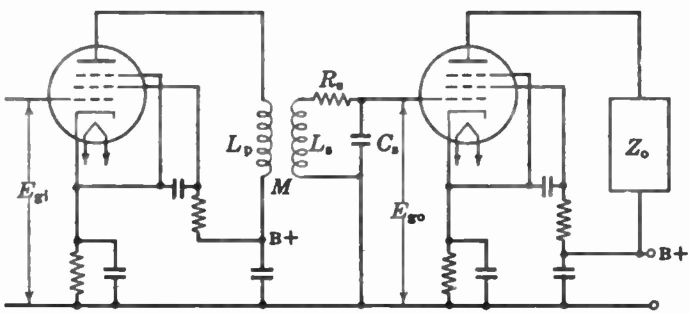
Because it is not practical to build coils that result in the quantity L/CR much higher than approximately 0.1 the plate resistance r_p of a pentode, the gain of an amplifier of this type is limited to about 0.1 the amplification factor of the tube. Also it is seen that if the Q_r factor of the coil remains constant over quite a range of frequencies the gain of the amplifier will be nearly directly proportional to the frequency to which it is tuned. This is not generally desirable, for example, in the tuned radio-frequency stage of a radio receiver.

---

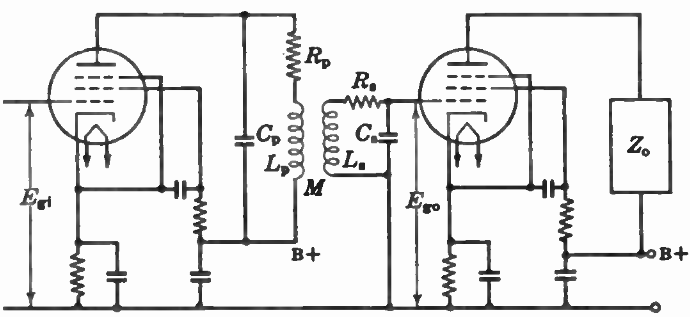
9. $\frac{I_{max}}{I} = \frac{f_r}{f}\sqrt{\frac{f^2}{f_r^2} + Q_r^2\left(\frac{f^2}{f_r^2} - 1\right)^2}$



(a) Single-tuned circuit.



(b) Tuned secondary transformer type.



(c) Double-tuned transformer type.

FIG. 3-45. Three types of tuned amplifiers.

Transformer Coupling With Tuned Secondary. The tuned secondary type of transformer coupling is shown in Fig. 3-45b. In the general case when both M and C_s can be adjusted for the maximum possible voltage across C_s these adjustments are derived in the following manner. The impedance that the circuit presents to the input tube is

$$\underline{Z}_o = R_p + \frac{R_s\omega^2M^2}{R_s^2 + X_s^2} + j\left(\omega L_p - \frac{X_s\omega^2M^2}{R_s^2 + X_s^2}\right) \qquad (3\text{-}75)$$

where $X_s = \omega L_s - \frac{1}{\omega C_s}$. Maximum power will be delivered to the secondary when $\underline{Z}_o - R_p = r_{p1}$ or $\underline{Z}_o = r_{p1} + R_p$ and, approximately, the voltage across C_s will be a maximum for any given voltage E_{g1}. This results in adjusting M and C_s so that

$$\omega L_p - \frac{X_s\omega^2M^2}{R_s^2 + X_s^2} = 0 \quad \text{and} \quad \frac{R_s\omega^2M^2}{R_s^2 + X_s^2} = r_{p1} + R_p \qquad (3\text{-}76)$$

For practically all cases R_p will be small compared to r_{p1} and can be neglected. Relations (3-75) reduce to the two following simultaneous expressions.

$$\frac{\omega L_p}{X_s} = \frac{r_{p1}}{R_s} \quad \text{and} \quad \omega M = \sqrt{r_pR_s + L_pX_s} \qquad (3\text{-}77)$$

These two relations can be satisfied simultaneously when

$$K^2 = \frac{1}{Q_{rs}}\left[\frac{1}{Q_{rp}} - \frac{1}{Q_{rs}} + Q_{rp}\right] \qquad (3\text{-}78)$$

where $Q_{rp} = \frac{\omega_r L_p}{r_{p1}}$, $Q_{rs} = \frac{\omega_r L_s}{R_s}$ and $K = \frac{M}{\sqrt{L_p\,L_s}}$.

When the above relations can be satisfied the voltage gain is

$$\frac{E_{go}}{E_{g1}} = \frac{\mu\,\frac{1}{\omega C_s}}{2\sqrt{r_{p1}\,R_s}} \qquad (3\text{-}79)$$

at a frequency $$f = \frac{f_{rs}}{\sqrt{1 - \frac{Q_{rp}}{Q_{rs}}}}$$

When the input tube is a pentode the practical value of Q_{rp} is so small that it would require K to be greater than one to satisfy the above expression. This is impossible. However, most present-day tuned amplifier circuits use pentode tubes and so it is necessary to have an expression for the voltage gain for this case.

When the constants of the input tube and circuit are such that the a-c plate current is substantially independent of the output impedance, i.e. r_{p1} is very large compared to Z_o, the voltage induced in the secondary circuit will be directly proportional to M and nearly independent of X_s, the net reactance of the secondary circuit.

$$\underline{I}_s = \frac{-j\omega M I_p}{R_s + jX_s} \tag{3-80}$$

Then I_p is nearly equal to $g_{gp1} E_{g1}$ and since

$$\underline{I}_s \cong \frac{-j\omega M\, g_{gp1} E_{g1}}{R_s + jX_s} \tag{3-81}$$

The voltage across the condenser C_s is $\underline{E}_{go} = +j \frac{1}{\omega C_s} \underline{I}_s$

Hence
$$\underline{A} \cong \frac{\underline{E}_{go}}{\underline{E}_{g1}} \cong \frac{\omega M\, g_{gp1}}{(R_s + jX_s)\omega C_s} \cong \frac{M/C_s\, g_{gp1}}{R_s + jX_s} \tag{3-82}$$

$$A \cong \frac{M/C_s\, g_{gp1}}{\sqrt{R_s^2 + X_s^2}} \tag{3-83}$$

A is a maximum when the frequency is adjusted to make $X_s = 0$ and substantially a maximum when the frequency is fixed and C_s is adjusted to make $X_s = 0$. In either case let this frequency be known as f_{rs}, the resonance frequency of the secondary circuit. Then

$$A \cong \frac{M/C_s\, g_{gp1}}{R_s} \cong \omega_r M \frac{\omega_r L_s}{R_s} g_{gp1} \tag{3-84}$$

$$\cong \omega_r M\, Q_{rs} g_{gp1} \cong \frac{\mu_1}{r_{p1}} Q_{rs} \omega_r M$$

where
$$Q_{rs} = \frac{\omega_r L_s}{R_s}$$

Since A depends upon frequency in the same manner as does the current in a series circuit of L, R, and C, the gain versus frequency characteristic of the amplifier are like the current versus frequency characteristic of a series resonant circuit. If Q_{rs} is substantially constant over a range of frequencies for which the amplifier is designed the maximum gain will be directly proportional to the operating frequency. This may be undesirable because the sensitivity of the system will increase with frequency.

When the input tube of an amplifier is a pentode the selectivity and the dependence of sensitivity on operating frequency are practically the same for the tuned plate circuit and the tuned-secondary transformer types of coupling. The maximum gain at the operating frequency can be made greater for the tuned-secondary transformer because M can be made larger than L of the tuned plate circuit type and have the same range of operation.

Doubly-tuned Transformer Coupling. The doubly-tuned transformer coupling type of radio-frequency amplifier is illustrated by Fig. 3-45c. This type of coupling is the one most generally used in broadcast receivers because, as will be seen later, it has a broader pass band and a sharper cutoff than either of the other two types, and for certain values of the circuit parameters the maximum gain is nearly constant over a range of frequencies.

In order to proceed with the analysis of the circuit of Fig. 3-45c the equivalent circuit is arranged as shown in Fig. 3-46. Then Thevenin's theorem[10] is used for arriving at the relations which must exist between the circuit constants for maximum voltage amplification.

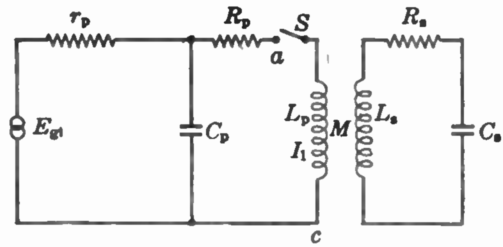


FIG. 3-46. Equivalent circuit of a double-tuned transformer-coupled amplifier.

With switch S of Fig. 3-46 open, the impedance $\underline{Z}_{bc}$ is

$$\underline{Z}_{bc} = \frac{R_s\omega^2M^2}{R_s^2 + X_s^2} + j\left(\omega L_p - \frac{X_s\omega^2M^2}{R_s^2 + X_s^2}\right) \tag{3-85}$$

The impedance $\underline{Z}_{ac}$, with the source voltage $(-\mu E_g) = 0$ is

$$\underline{Z}_{ac} = R_p + \frac{r_p}{1 + r_p^2\omega^2C_p^2} - j\,\frac{r_p^2\omega C_p}{1 + r_p^2\omega^2C_p^2} \tag{3-86}$$

For maximum transfer of power into the secondary circuit $\underline{Z}_{bc}$ must be the conjugate of $\underline{Z}_{ac}$. The voltage across condenser C_s will be substantially a maximum when the power delivered to the secondary circuit is a maximum even though f or C_s are adjusted to meet the power conditions. Making $\underline{Z}_{bc}$ equal to the conjugate of $\underline{Z}_{ac}$ results in the simultaneous relations which must be satisfied for maximum voltage amplification namely,

$$R_p + \frac{r_p}{1 + r_p^2\omega^2C_p^2} - \frac{R_s\omega^2M^2}{R_s^2 + X_s^2} = 0 \tag{3-87}$$

$$\omega L_p - \frac{X_s\omega^2M^2}{R_s^2 + X_s^2} - \frac{r_p^2\omega C_p}{1 + r_p^2\omega^2C_p^2} = 0 \tag{3-88}$$

Only one special case will be analyzed. This is the case when a pentode is used for the input tube, so that $r_p^2 \gg \frac{1}{\omega^2C_p^2}$ for all practical values of C_p. Under these conditions the two simultaneous relations reduce to

---

10. Thevenin's Theorem is given in Ch. 1.

$$\omega^2M^2 = \frac{\left(\frac{1}{r_p\omega^2C_p^2} + R_p\right)\left(R_s^2 + X_s^2\right)}{R_s} \tag{3-89a}$$

$$\frac{\frac{1}{r_p\omega^2C_p^2} + R_p}{R_s} = \frac{X_p}{X_s} \tag{3-89b}$$

where $X_p = \omega L_p - \frac{1}{\omega C_p}$ and $X_s = \omega L_s - \frac{1}{\omega C_s}$

The quantity $\frac{1}{r_p\omega^2C_p^2}$ may be regarded as effectively increasing the resistance of the primary circuit. Then if the Q factors of the two circuits are defined as follows

$$Q_{rp} = \frac{\omega_{rp}L_p}{\frac{1}{r_p\omega^2C_p^2} + R_p} \quad \text{and} \quad Q_{rs} = \frac{\omega_{rs}L_s}{R_s} \tag{3-90}$$

the simultaneous relations for maximum voltage gain can be rearranged into the two forms.

$$\frac{Q_{rs}}{Q_{rp}} = \frac{1 - \frac{f_{rp}^2}{f^2}}{1 - \frac{f_{rs}^2}{f^2}} \tag{3-91}$$

and

$$k^2 = \frac{\omega_{rp}\omega_{rs}}{\omega^2Q_{rp}Q_{rs}} + \left(1 - \frac{f_{rp}^2}{f^2}\right)\left(1 - \frac{f_{rs}^2}{f^2}\right) \tag{3-92}$$

where $k = M/\sqrt{L_pL_s}$, $f_{rp} = \frac{1}{2\pi\sqrt{L_pC_p}}$ and $f_{rs} = \frac{1}{2\pi\sqrt{L_sC_s}}$

The solution of these two expressions yield the following expressions at which maximum voltage gain occurs,

$$f_m \cong \sqrt{\frac{-(f_{rp}^2 + f_{rs}^2) \pm (f_{rp}^2 + f_{rs}^2)^2 + 4f_{rp}^2f_{rs}^2(k^2 - 1)}{2(k^2 - 1)}} \tag{3-93}$$

and

$$f_m \cong \sqrt{\frac{Q_{rs}f_{rs}^2 - Q_{rp}f_{rp}^2}{Q_{r2} - Q_{r1}}} \tag{3-94}$$

when

$$\frac{f_{rp}f_{rs}}{Q_{rp}Q_{rs}} \ll (f_{rs}^2 + f_{rp}^2).$$

Expression (3-91) shows that when f_{rp} is not equal to f_{rs} there can be only one frequency at which the voltage amplification can reach the optimum value and this frequency must also satisfy expression (3-92). In general the voltage amplification versus frequency characteristic will have two peaks, when the coupling is greater than a certain critical value, but when f_{rp} is not equal to f_{rs} one of the peaks is higher than the other and neither may be optimum, unless the circuit constants satisfy both of the expressions for f_m.

When the two circuits are adjusted so that $L_s C_s = L_p C_p$ and $Q_{rp} = Q_{rs}$ then the expression (3-91) is satisfied for all frequencies. This is nearly true also when Q_{rs} is not quite equal to Q_{rp}. The expression for f_m becomes

$$f_m \cong \frac{f_r}{\sqrt{1 \pm k}} \tag{3-95}$$

or

$$f_{m1} \cong \frac{f_r}{\sqrt{1 + k}} \quad \text{and} \quad f_{m2} \cong \frac{f_r}{\sqrt{1 - k}}$$

where $f_r = f_{rp} = f_{rs}$ and the coupling is greater than critical. This results in two values of equal magnitude in the gain versus frequency characteristic and both of these values are optimum. The critical or smallest value of coupling the system can have and still produce the optimum gain is $k^2 = \frac{1}{Q_{rp} Q_{rs}}$ which results, as seen from expression (3-92), when $f = f_r$. Then for critical coupling

$$f_{m1} \cong f_r \quad \text{and} \quad f_{m2} \cong \frac{f_r}{\sqrt{1 - k^2}} \tag{3-96}$$

These two frequencies are usually so close together that there is no appreciable drop in the voltage gain between them. This is the reason for the flat top nature of the gain versus frequency characteristic. It is necessary to go back to expression (3-92) to arrive at the expression for f_{m1} and f_{m2} in (3-96) because when $k = 1/Q_{rp}Q_{rs}$ the approximation which was used in getting f_{m1} and f_{m2} for coupling greater than critical is not valid. Figure 3-47 illustrates a typical gain versus frequency

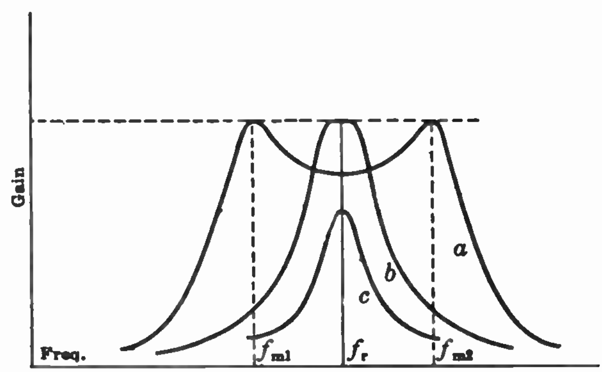


FIG. 3-47. Typical gain vs frequency characteristics of a double-tuned transformer type of tuned amplifiers. $L_p C_p = L_s C_s$ and $Q_{rp} = Q_{rs}$
(a) Coupling greater than critical.
(b) Critical coupling.
(c) Coupling less than critical.

characteristic for the type of circuit in which $f_{rp} = f_{rs} = f_r$ and $Q_{rp} = Q_{rs}$.

The optimum gain of the amplifier stage is deduced from the power relation in the following manner. Referring to Fig. 3-46 the circuit is adjusted so that $\underline{Z}_{bc}$ is the conjugate of $\underline{Z}_{ac}$.

Hence
$$I_s^2 R_2 = I_1^2\left(R_p + \frac{r_p}{1 + r_p\omega^2 C_p^2}\right)$$

but
$$I_1 = \frac{E_{ac(\text{with S open})}}{2\left(R_p + \frac{r_p}{1 + r_p\omega^2\omega^2 C_p^2}\right)}$$

$$= \frac{\mu E_{g1}}{2\left(R_p + \frac{r_p}{1 + r_p\omega^2 C_p^2}\right)\sqrt{1 + r_p^2\omega^2 C_p^2}}$$

The voltage $E_{go} = I_s \frac{I}{\omega C_s}$

From these expressions there results optimum gain

$$A = \frac{\mu\frac{1}{\omega C_s}}{2\sqrt{[R_p(1 + r_p^2\omega^2 C_p^2) + r_p]R_s}} \qquad (3\text{-}97)$$

This expression shows that the optimum gain is slightly less than that for the tuned-secondary type because of the quantity $R_p(1 + r_p^2\omega^2 C_p^2)$ in place of R_p. However the doubly-tuned type is superior from a selectivity standpoint because the gain over the pass band is more uniform and the change from passing to attenuating is more abrupt.

The doubly-tuned transformer coupling is used mostly in the intermediate-frequency amplifier of a superheterodyne receiver where the tuning is adjusted for a constant band width and operating frequency and it is not necessary to make the two circuits "track" together over a range of frequencies.

3-19. Video-Frequency Amplifiers.- A video-frequency amplifier is one that is used for amplifying television picture or other impulses to a level necessary for modulating a transmitter, or the rectified signals at the receiver to a level necessary for operating a cathode ray tube. The electrical impulses produced from the picture pick-up tube or other source contain a very wide band of frequencies,[11] much wider than is

11. In television, assuming the scene to be transmitted is "divided" into 441 lines and scanned 30 times per second, the scanning of one line takes 1/13,230 sec. If there are 50 evenly spaced small black rectangles on a line, the change from black to white and back to black in scanning the line is at a rate of 50 x 13,230 = 651,500 per second. To transmit up to the third harmonic of this frequency would require a 2-megacycle band.

(Footnote continued)

encountered in audio-frequency work. These frequencies lie in a spectrum from a few cycles per second to a few megacycles per second; in some cases the upper limit is above 10 megacycles per second. Unlike audio-frequency amplifiers applied to electro-acoustical systems, the video-frequency amplifier must be free from phase distortion. Phase distortion cannot be tolerated because it results in a picture image with distorted detail, or in a reproduced pulse which does not resemble the original closely.

A. RESISTANCE-CAPACITANCE COUPLING WITH COMPENSATION. To design a video-frequency amplifier, with tubes very much like those used for audio-frequency amplifiers, for nearly constant gain and little phase distortion from a few cycles per second to a few megacycles per second requires much more judgment and care than are required in the design of an audio-frequency amplifier. It will be remembered that high gain and a gain that is constant with frequency for the higher frequencies are incompatible. Therefore to extend the upper limit of the range of an amplifier to several megacycles or more requires a great reduction in gain from the value that can be obtained in an audio-frequency amplifier using the same tubes. For an example, suppose the input tube of a resistance-capacitance-coupled amplifier is a pentode, $r_p = 0.75 \times 10^6$ megohms, $g_{gp} = 9000$ micromhos and $\mu = 6750$ and assume that the capacitances, including sockets and wiring, of the input tube and output tube are such as to impose a capacitance of 20×10^{-12} farads across the coupling resistors. The susceptance of 20×10^{-12} farads at 10^6 cycles is 125 micromhos. For a reduction of 1 Db below the medium frequency gain the conductances of the coupling resistor plus the plate conductance would have to be 277 micromhos. The gain would be $g_{gp}/277 \times 10^{-6} = 32.5$ and the combined resistance of the two coupling resistors would be 3630 ohms. The phase shift would be approximately 27° which would be very bad for a video-frequency amplifier. Furthermore one megacycle is not high enough for a video-frequency amplifier in a good television transmitter or receiver.

High-frequency Compensation. Some compensation of gain and phase shift can be achieved at the high frequencies in a resistance-capacitance-coupled video-frequency amplifier by placing a small inductance in series with the coupling resistor

---

Footnote continued

Pulses of a microsecond duration or less are in practical use; to amplify such a pulse would require an amplifier to pass a fundamental frequency of one megacycle and such harmonic frequencies as may be required to give specified reproduction of the original.

R_c. The circuit diagram is shown in Fig. 3-48. The expression for the voltage amplification at the high frequencies is

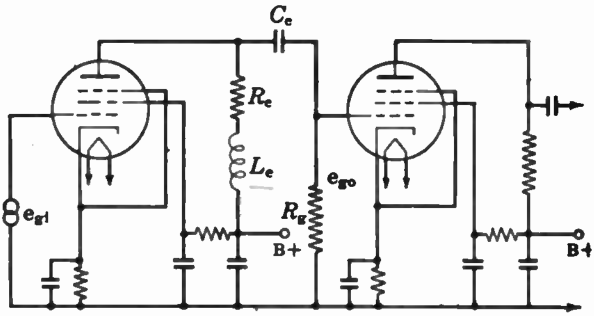


FIG. 3-48. Resistance-capacitance-coupled amplifiers with high-frequency compensation by L_c.

$$A_H = \frac{E_{go}}{E_{gi}} = \frac{-g_{gpi}}{g_{pi} + G_{go} + \frac{R_c}{R_c^2 + \omega^2 L_c^2} + j\left(\omega C_e - \frac{\omega L_c}{R_c^2 + \omega^2 L_c^2}\right)} \qquad (3\text{-}98)$$

where $\frac{R_c}{R_c^2 + \omega^2 L_c^2} - j\,\frac{\omega L_c}{R_c^2 + \omega^2 L_c^2}$ is the admittance of the coupling resistor circuit and C_e is the effective capacitance shunting the coupling circuit. For a pentode C_o is nearly equal to the output capacitance of the input tube plus the input capacitance of the output tube, including sockets and wiring.

Pentodes are used for video-frequency amplifiers and the frequency range extends into a few megacycles. Then the constants of the circuit are such that g_{pi} and G_{go} in the above expressions are negligible compared to the other quantities of the denominator. Under these conditions

$$A_H \cong \frac{g_{gp} R_c \sqrt{\frac{f_e^2}{f^2} + D^2}}{\sqrt{\frac{f_e^2}{f^2} + \frac{f^2}{f_e^2} D^2 + 1 - 2D}} \qquad (3\text{-}99)$$

where $D = \frac{L_c}{C_e R_c^2}$ and $2\pi f_e = \frac{1}{R_c C_e}$

Setting $2\pi f_e = \frac{1}{R_c C_e}$ makes $\frac{f}{f_e} = 1$ when the reactance of the capacitance C_e is equal to the coupling resistance R_c. This also permits changing D by changing L_c without disturbing f_e. Without the compensation due to L_c, i.e. $D = 0$ because $L_c = 0$,

$$\underline{A}_H \cong \frac{-g_{gpi}R_c}{1 + jR_c\omega C_e} \cong \frac{-g_{gpi}R_c}{1 + j\,\dfrac{f}{f_e}} \qquad (3\text{-}100)$$

$$A_H \cong \frac{g_{gpi}R_c}{\sqrt{1 + R_c{}^2\omega^2 C_e{}^2}} = \frac{A_M}{\sqrt{1 + \dfrac{f^2}{f^2}}} \qquad (3\text{-}101)$$

where $A_M = -g_{gpi}R_c$ is the medium-frequency gain.
With compensation the phase shift is

$$\theta'_L \cong \tan^{-1}\left(\frac{f}{f_e} + \frac{f^3}{f_e{}^3}D^2 - \frac{f}{f_e}D\right) \qquad (3\text{-}102)$$

Without compensation the phase shift is $\theta_L \cong \tan^{-1} f/f_e$.

Figure 3-49 shows how gain and phase shift are related to various values of D, including the uncompensated case, D = 0. A

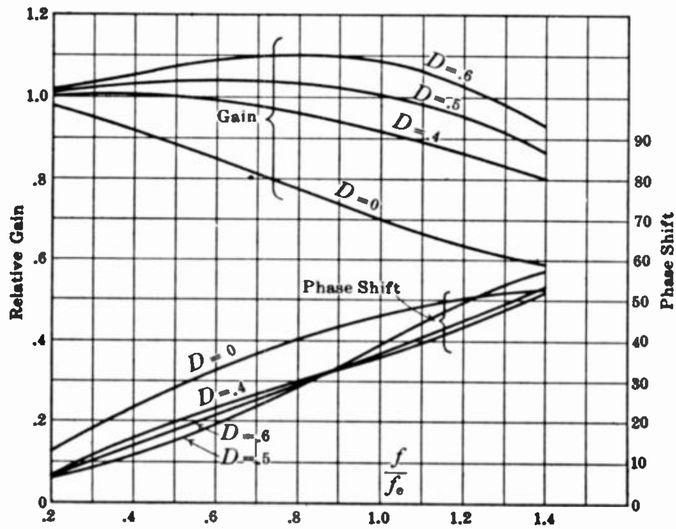


FIG. 3-49. Gain and phase shift characteristics of a video-frequency amplifier with high-frequency compensation.

value of D between 0.4 and 0.5 gives almost constant gain up to a frequency equal to f_e but there is not much reduction in phase shift from the uncompensated case when $f = f_e$. Since the capacitance C_e is fixed by the kind of tubes, sockets and wiring there is no control over C_e after a selection of these has been made. Thus the only way to achieve a high value for f_e is to select R_c properly and the higher f_e is made the lower will be the gain of the amplifier over the entire range. After selecting

R_c for the predetermined frequency f_e an inductance L_c is added in series with R_c which will give a value of D between 0.4 and 0.5. It should be recognized that the compensated amplifier has a higher medium-frequency gain than the uncompensated type for the same gain versus frequency characteristics.

Low-frequency Compensation. It was pointed out in Sec. 3-6 on the resistance-capacitance coupled amplifier how low-frequency compensation can be achieved. In the case of video-frequency amplifiers where special pentodes are used the approximations involved in arriving at the simpler expression for the gain are valid. That is $R_s \gg 1/\omega C_s$, $R_{go} \gg R_c$ and $R_c + 1/(\omega^2 C_s^2 R_s) \ll r_p$. Then the gain and phase shift are given by expressions 3-24 and 3-25.

B. LOW-PASS FILTER COUPLING. Since it is desirable to have a type of coupling impedance that is constant and either produces no phase shift or has a phase shift that is directly proportional to the frequency for a video amplifier it occurred to some of the early workers in this field that the proper kind of a low-pass filter could be used in place of the coupling resistance R_c. When a low-pass filter is terminated in its characteristic impedance its input impedance is equal to the characteristic impedance. Certain m-derived filters have characteristic impedances that are nearly constant and nearly equal to a pure resistance over the pass band. Hence if such a filter can be found, nearly matching a pure resistance, the filter can be terminated in a resistance. The shunt m-derived T type has nearly these properties when m = 0.6. Then for the coupling impedance a low-pass filter is built up which makes use of the tube capacitances for the shunt arm capacitances and the terminating section is a half T section of shunt m-derived. A typical filter of this type is shown in Fig. 3-50. The input impedance of this filter is nearly equal to the characteristic impedance of a π type constant-K section. The characteristic impedance of the π type constant-K filter is, in terms of the cutoff frequency f_c,

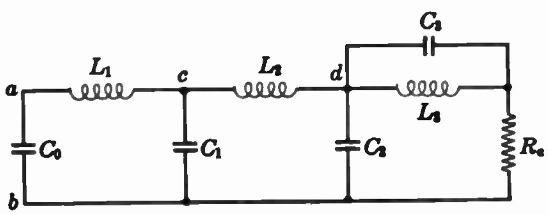


FIG. 3-50. Schematic low-pass filter for an amplifier coupling impedance.

$$Z_{c\pi} = \frac{R_c}{\sqrt{1 - (\frac{f}{f_c})^2}} \quad (3\text{-}103)$$

This is approximately the input impedance, measured across the terminals a-b, of the filter. It is desired to have the input impedance constant over the pass band so that the voltage amplification of the input tube will remain constant. This can be accomplished by adding capacitance in parallel with C_o. When this is done the expression for the input impedance of the filter becomes

$$\underline{Z}_1 = \frac{1}{\frac{1}{\underline{Z}_{c\pi}} + j\omega C_s} \quad (3\text{-}104)$$

where C_s is the added shunt capacitance. Substituting in the above expression the expression for $\underline{Z}_{c\pi}$ there results

$$\underline{Z}_1 = \frac{R_c}{\sqrt{1 - (f/f_c)^2} + j\omega C_s R_c} \quad (3\text{-}105)$$

$$Z_1 = \frac{R_c}{\sqrt{1 - (f/f_c)^2 + \omega^2 C_s^2 R_c^2}} \quad (3\text{-}106)$$

The input impedance will be independent of f when

$$\omega^2 C_s^2 R_c^2 = \frac{\omega^2}{\omega_c^2} \quad \text{or} \quad C_s = \frac{1}{\omega_c R_c} = \frac{1}{2\pi f_c R_c}$$

$1/\pi f_c R_c$ is the total shunt capacitance of a constant-K low-pass filter. Hence the input impedance to the filter of Fig. 3-50 will be substantially constant over the pass band when the capacitance C_o is equal to the total shunt capacitance of a constant-K filter and when the m-derived half T section next to the load resistance R_c has a value of $m = 0.6$.

The phase shift through an ideal amplifier should either be zero or directly proportional to the frequency. Under these conditions the velocity of transmission will be independent of frequency and the wave shape of the output will be exactly the same as the input. When no capacitance is added to the input of the filter, which is used for a coupling impedance, the phase angle of the input impedance is zero and the output voltage of the input tube is 180° out of phase with the input voltage. However the gain varies with the frequency. When shunt capacitance equal to $1/2\pi f_c R_c$ is added to the input of the filter the gain is independent of the frequency and the phase shift through the tube becomes lagging and is given by the expression

$$\theta = \tan^{-1} \frac{1}{\sqrt{\frac{f_c^2}{f^2} - 1}} \quad (3\text{-}107)$$

The phase shift varies from 0 to 90° over the pass band and is not quite linear with frequency. The departure from linearity occurs mostly just below the cutoff frequency. When the added capacitance is made a little greater than the above value the phase shift becomes nearly linear with frequency but the impedance varies with frequency. Figure 3-51 illustrates the phase shift and gain. The gain over the pass band is $A = g_{gp}R_c$. The constants of a filter of this type, Fig. 3-50, are related to each other as follows:

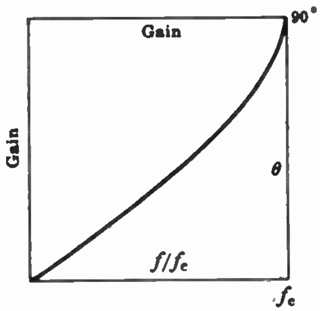


FIG. 3-51. Phase shift when $C_o = 2/\pi f_c R_c$.

$$L_1 = L_2 = \frac{R_c}{\pi f_c}, \quad C_o = C_1 = \frac{1}{\pi f_c R_c}$$

$$C_2 = 0.5\,\frac{1}{\pi f_c R_c} + m\,\frac{0.5}{\pi f_c R_c}$$

$$C_3 = \frac{1 - m^2}{4m}\;\frac{2}{\pi f_c R_c} \quad \text{and} \quad L_3 = \frac{mL_1}{2}$$

When $m = 0.6$, C_2 is 80 per cent of $\frac{1}{\pi f_c R_c}$

If the plate of the input tube and the grid of the output tube were both connected to the input terminals a-b of the filter, C_o would be the output and input capacitances of these tubes and this would be the lowest possible value of C_o and the maximum possible gain would be $g_{gp}/\pi f_c C_o$. Since there is no attenuation over the pass band the voltage developed between points c-b and d-b is the same as the voltage a-b and differs only by the phase displacement. Consequently the plate of the input tube may be connected across the

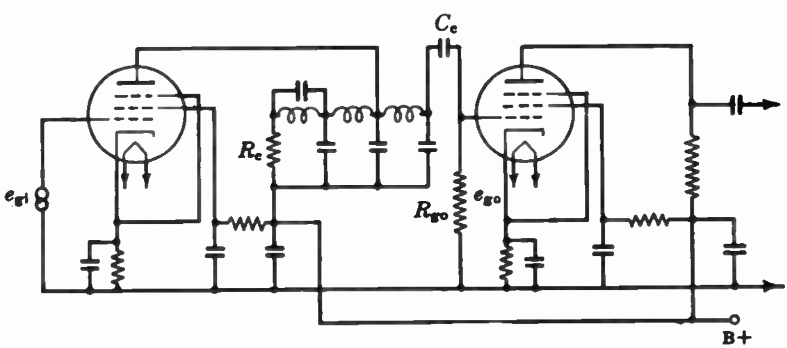


FIG. 3-52. Low-pass filter arrangement for a coupling impedance in a video-frequency amplifier.

input terminals a-b and the grid of the output tube connected to terminals c-b, or c-d, and vice versa. This connection separates the tube capacitances and permits reducing the shunt capacitances of the low-pass filter and thereby makes it possible to obtain a higher gain.

In order to keep the d-c plate voltage off the grid of the output tube it is necessary to use a coupling condenser and grid leak resistor. This combination introduces amplitude and phase distortion at the lowest frequencies which can be compensated by the method given for a resistance-capacitance-coupled amplifier.

In video-frequency amplifier construction more care must be taken than with an audio-frequency amplifier because of the high-frequency range of the amplifier. By-pass condensers for the screen and plate circuits must be large to be effective at the lowest frequency and must be paralleled with small r-f condensers to be effective at the highest frequencies. Cathode by-pass condensers must be very large in order to prevent cathode current feed-back at the low frequencies. All plate and grid leads of the tubes should be as short as possible and the grid coupling condenser should have as small a capacitance to ground as possible. The coupling resistors and filter inductances should not have excessive residual capacitances. The coupling resistance of the resistance-capacitance-coupled amplifier can be inductive provided its inductance is equal to or less than the inductance required with a pure resistance to give the proper compensation at the highest frequencies.

PART 4. RADIO FREQUENCY POWER AMPLIFIERS

3-20. The Class C Power Amplifier.- When the output circuit of an amplifier is of such nature that the impedance to the operating frequency is quite high and has nearly unity power factor whereas to all other harmonic frequencies the impedance is very low and has nearly zero power factor, operation of a single tube need not be confined to its nearly linear volt-ampere characteristics. With an output impedance of this nature the tube generates, or converts, a large amount of power of the fundamental operating frequency as compared to the harmonic power even though the a-c plate current is made up of a fundamental and a large amount of harmonics. Figure 3-53 illustrates a typical Class C amplifier circuit except no neutralization of the grid-to-plate capacitance is shown. For a given d-c plate potential the grid is biased beyond cutoff and a-c plate current flows for only a fraction of the input cycle. An amplifier operated in this manner is known

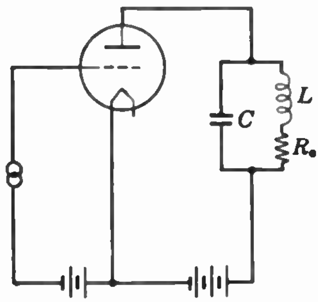


FIG. 3-53. Schematic diagram of Class C amplifier circuit; neutralization not shown.

as a Class C power amplifier.[12] Figure 3-54 illustrates the mode of operation and the general nature of the a-c plate current and a-c plate voltage. The a-c plate voltage is nearly a sine wave because the impedance of the parallel circuits of L, R_e and C is low to all harmonic frequencies except the fundamental. The efficiency of power conversion is relatively high because practically all of the plate current flows when the instantaneous plate potential is considerably below the average plate potential. This results in a given a-c power output for less d-c power input than is found in the other classes of amplifiers.

Analysis of a Class C Amplifier. Both the mathematical and graphical analysis of a Class C amplifier are much more involved than that of a Class A amplifier. In the Class A amplifier operation is confined to the nearly linear characteristics

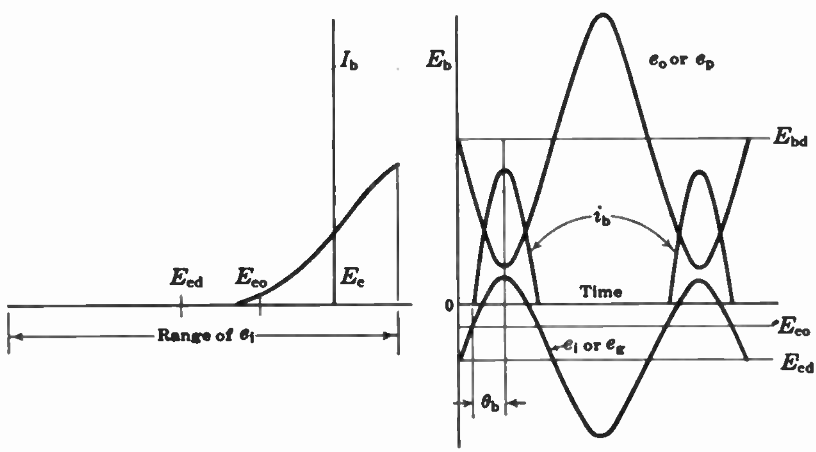


FIG. 3-54. Mode of operation of a Class C amplifier. θ_b is angle of plate current flow.

of the tube where μ and r_p are essentially constant over the complete cycle of the input voltage and the analysis becomes

12. See Sec. 3-1.

one of finding the maximum ranges of the a-c plate potential and current for a given set of operating conditions that will result in nearly sinusoidal operation. Then the power output and efficiency are determined from rather simple expressions. In the Class C amplifier, for a given d-c plate potential, the grid bias can be almost any value between cutoff and about three times cutoff. The angle of the input voltage over which plate current flows depends upon the d-c grid potential and the a-c grid and plate potentials. The peak value of the plate current depends upon the volt-ampere characteristics of the tube, the load impedance and the a-c input potential. The fundamental component of the plate current depends upon the angle of flow and the peak value of the total plate current. Consequently for every selected set of d-c grid and plate potentials there are many variables and any mathematical analysis must start with some assumed relation between plate current and grid and plate potentials such as:[13]

$$i_b = g_{gp}\left(e_c + \frac{e_b}{\mu}\right) \text{ for } e_c + \frac{e_b}{\mu} > 0$$

and $$i_b = 0 \text{ for } e_c + \frac{e_b}{\mu} < 0$$

or[14]

$$\text{Total Space Current} = i_b + i_c = K\left(e_c + \frac{e_p}{\mu}\right)^{3/2}$$

For the mathematical analysis based on these two assumptions the reader is referred to the two references given at the bottom of the page.

The other alternative is to assume d-c operating potentials and work the problem graphically from the constant plate current E_c - E_b characteristics or the constant grid voltage I_b - E_b characteristics. This too requires a large amount of work to arrive at the a-c input voltage, tank circuit impedance, and finally the power output for each set of assumed d-c potentials. There are however some considerations that have been obtained from practice that reduce the amount of work. In the first place the plate current is near enough to a part of sine wave that it can be considered thus. On this basis curves of i_{bm}/I_{bd} and i_{bm}/I_p plotted against θ_b, the angle of plate current flow, are useful and are shown in Fig. 3-55; i_{bm} is the maximum instantaneous plate current, I_{bd} is the d-c component of plate current and I_p is the peak value of the fundamental component of plate current. The angle θ_b is measured from the peak to the cutoff as shown[15] in Fig. 3-54. This same figure

13. "Communication Engineering," Everitt.

14. "Radio Engineering," Terman.

15. Wagener, "Performance of Transmitting Tubes," Proc. I. R. E. Jan. 1937.

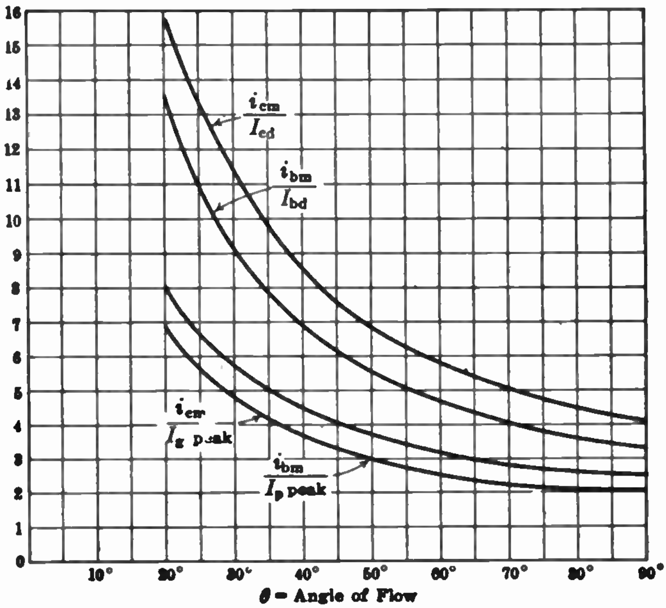


FIG. 3-55. Curves of current ratios in a Class C amplifier. Taken from Wagener. I_g and I_p are peak values of the fundamental.

also shows the corresponding grid current relation which will be discussed later.

Usually the tube manufacturer gives data on the maximum d-c plate voltage, plate dissipation and a-c power output for each Class C power amplifier. Assuming then a power output of P_o watts equal to or less than the manufacturer's rating, the maximum value of the plate current i_{bm} times the peak value of the a-c plate potential E_p is given by the relation,

$$E_p i_{bm} = 2D_1 P_o \qquad (3\text{-}108)$$

where $D_1 = \frac{i_{bm}}{I_p}$ for an angle of flow of θ_1 degrees, and I_p and E_p are the peak values of the fundamental components of a-c plate current and potential respectively. Next assume an angle θ_1 of current flow. From the curves of Fig. 3-55 it is seen that for angles of flow between 60° and 90° D does not change rapidly. Hence any reasonable value such as 70° will not be far from the best operating conditions. With an assumed angle of current flow the quantity $E_p i_{bm}$ becomes fixed. Now locate this product on each of the constant current curves of the E_c - E_b characteristics. For each constant current value I_b' the voltage E_b' will have to be

$$E_b' = E_{bd} - E_p = E_{bd} - \frac{2D_1 P_o}{I_b'} \qquad (3\text{-}109)$$

where E_{bd} is the d-c operating plate potential and E_b' is the instantaneous plate potential for the current I_b'. Greater accuracy will be obtained if the constant current lines are projected as straight lines into the region of low plate potentials and the point I_b' - E_b' is located on its corresponding straight line instead of the actual curve which curves upward badly when the plate potential is low and the grid potential high. Only a few of the products of $E_b i_{bm}$ will have to be located and these for the higher values of plate current. Now draw a curve through these points. From this curve it becomes apparent what the lower limit of the plate current i_{bm} must be if the grid potential is

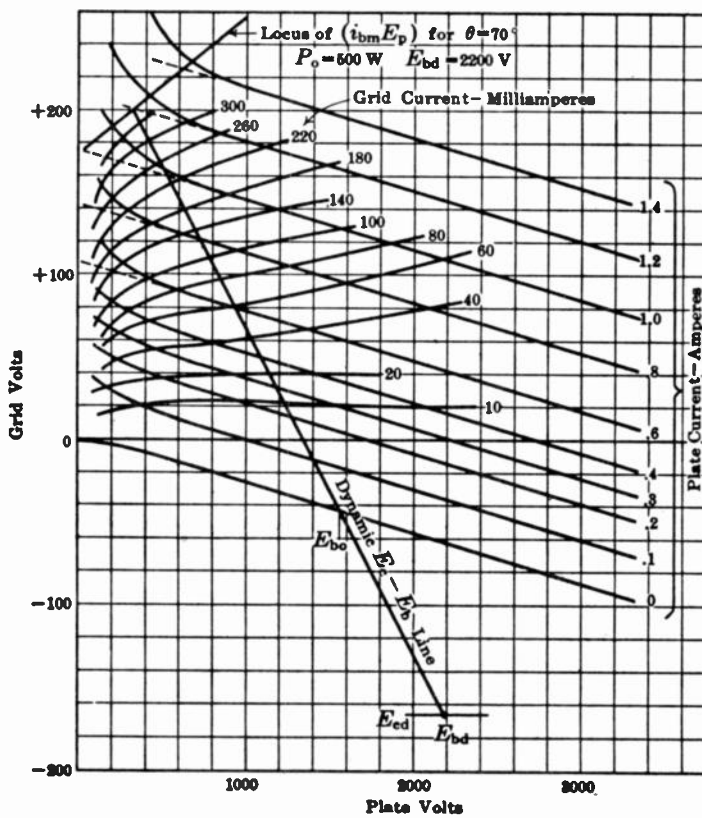


FIG. 3-56. Graphical solution of a Class C amplifier from $E_c - E_b$ characteristic.

to be a little less than the minimum plate potential and still have the tube deliver the assumed power output with the assumed angle of flow. This is illustrated in Fig. 3-56.

The next step is to draw a line from the tentatively selected i_{bm} point, which lies on the locus of the $E_p i_{bm}$ values, to the d-c operating plate voltage such that it crosses the zero constant current line at a plate voltage E_{bc} given by the relation

$$E_{bc} = E_{bd} - E_p \cos \theta.$$

Where this line, which is the dynamic voltage line, crosses the operating d-c voltage line E_{bd} will determine the grid bias voltage E_{cd} needed for the assumed conditions. The absolute sum of the grid voltages lying between the two extremities of the dynamic voltage line is the peak a-c driving voltage required to produce the maximum value of the instantaneous plate current and to produce the assumed power output.

The d-c power input is found by getting the ratio of i_{bm} to I_{bd} for the assumed angle of flow from the curve of Fig. 3-55. Calling this ratio F_1, the d-c power input becomes

$$P_{dc} = \frac{i_{bm}}{F_1} E_{bd}. \qquad (3\text{-}110)$$

The plate loss is $P_p = P_{dc} - P_{ac}$

The plate circuit efficiency is P_{ac}/P_{dc}

The tank circuit impedance R_{01} required to produce the assumed operation is

$$R_{01} = \frac{E_p^2}{2P_{ac}} \qquad (3\text{-}111)$$

where E_p is the peak value of the a-c voltage required to produce the maximum current i_{bm}.

If the results of the first trial show too much plate dissipation, or insufficient power output, or too low efficiency, then a new value of i_{bm} or θ must be assumed and the procedure carried out as before. If many calculations have to be carried out it will save time in locating the grid bias to be used by constructing a triangle on transparent material like the one shown in Fig. 3-57. This triangle is used by placing its base line a-b over the constant potential line E_{bd}. The minimum plate potential line $E_{bd} - E_p$ will then cross the slant side a-c at a point p. A line parallel to the side b-c through point p will cross the flow angle line at a potential $E_{\theta 1}$ which will be the potential at which the dynamic $E_c - E_b$ cuts the constant cutoff current line.

<u>Grid Circuit Relations</u>. One approximate method for getting the a-c driving power and the power loss in the grid

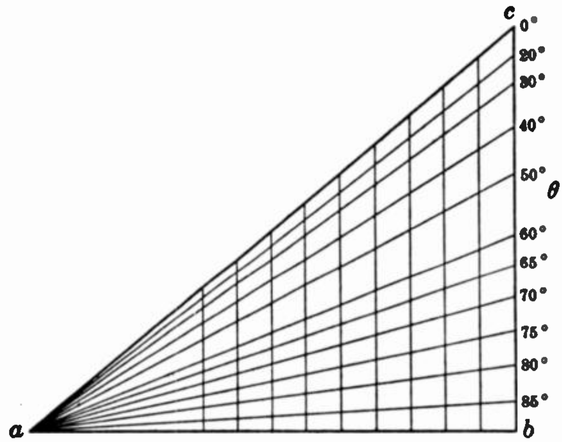


FIG. 3-57. Triangle for facilitating the location of a dynamic line on $E_c - E_b$ chart.

assumes the grid current is part of a sine squared function. The curves for the ratio of maximum grid current to the d-c grid current and maximum grid current to peak value of the fundamental component are used in the same manner as the corresponding curves for the plate circuit. To carry out the analysis by this method involves the angle of grid current flow which is considerably less than the angle of plate current flow. The angle of grid current flow depends only on the grid bias potential and the maximum positive grid potential because grid current flows only while the grid is positive with respect to the cathode. The maximum value of the grid current is determined from the grid current characteristic and the peak value of the fundamental component of grid current and d-c grid current are determined from the curves of Fig. 3-55.

The a-c driving power is $I_{g1}E_g/2$. Part of this power is dissipated in the C bias and the rest is dissipated by the grid. The power dissipated in the C bias is $E_{cd}I_{cd}$. Hence the power dissipated by the grid is

$$P_g = \frac{I_{g1}E_g}{2} - E_{cd}I_{cd}. \qquad (3\text{-}112)$$

where I_{g1} and E_g are peak values.

Tank Circuit Considerations. The impedance presented to the tube by the parallel circuit of L, R, and C and the load is

$$\underline{Z}_o = \frac{R_e X_c^2}{R_e^2 + X_n^2} - j\,\frac{R_e^2 X_c + \frac{L}{C} X_n}{R_e^2 + X_n^2} \qquad (3\text{-}113)$$

where $X_c = \frac{1}{\omega C}$, $X_L = \omega L$ and $X_n = \omega L - \frac{1}{\omega C}$ and R_e is the resistance of coil L plus the load resistance transferred effectively in series with L. If the load is coupled or connected to L in such a manner that it changes the reactance of the coil branch this also must be taken into consideration.

For the operating frequency f, the circuit is adjusted to have an impedance substantially, $\underline{Z}_{o1} = \frac{\omega_1^2 L^2}{R_e} = \omega_1 L Q_{e1}$ which is resistance and has unity power factor. This is the impedance R_{o1} that was previously used for the relation between a-c plate potential and a-c power output. The fundamental component of plate current flowing through this impedance produces the a-c plate voltage. The a-c plate voltage will be nearly a sine wave if the impedance Z_o is small at the harmonic frequencies. The impedance $\underline{Z}_o$ for the second harmonic frequency is approximately

$$\underline{Z}_{o2} = \frac{R_e}{9} - j\,\frac{2}{3}\,\omega_1 L$$

which is largely reactive. Neglecting the resistive term and using Z_{o1} and Q_{e1}

$$\underline{Z}_{o2} = -j\,\frac{Z_{o1}}{1.5 Q_{e1}} \text{ where } Q_{e1} \text{ is } \frac{\omega_1 L}{R_e} \tag{3-114}$$

For the third harmonic

$$\underline{Z}_{o3} = -j\,\frac{Z_{o1}}{2.75 Q_{e1}} \tag{3-115}$$

For Q_{e1} equal to greater than 10 the impedance for the second harmonic is equal to or less than 6.66 per cent of the fundamental frequency impedance, and the third harmonic impedance is equal to or less than 3.6 per cent of the fundamental frequency impedance. The second and third harmonic currents are much smaller than the fundamental component. Hence with a Q_{e1} factor greater than 10 the harmonic content in the plate voltage is only a few per cent of the fundamental.

In practice it is generally easy to measure the a-c current in the tank coil or condenser circuit and compute the a-c plate voltage. The a-c plate voltage times the coil current gives the volt-amperes in the coil circuit. From these volt-amperes and the watts output Q_{e1} is obtained as follows

$$Q_{e1} = \frac{\omega_1 L}{R_e} = \frac{\omega_1 L I_L^2}{R_e I_L^2} = \frac{E I_L}{P_o} \tag{3-116}$$

where $I_L (= E/\omega L)$ is the current in the coil branch. The current in the coil branch is nearly equal to the current in the condenser branch.

Grid Bias. Grid bias for a Class C amplifier is usually obtained by placing a resistor in the grid circuit. This resistor is by-passed by a condenser so that the impedance is low to the driver. The d-c grid current flowing through the bias resistor gives the necessary bias voltage for the grid. Hence for a bias voltage of E_{cd} volts the value of the grid resistor must be $R_{gb} = E_{cd}/I_{cd}$. For cases where the exciting voltage is apt to be interrupted some fixed bias is necessary to prevent damage of the tube because the self-bias will be zero and the tube will draw a heavy d-c plate current unless some fixed bias is provided.

3-21. Class B Tuned Power Amplifier.- The Class B tuned power amplifier can be considered as a special case of Class C operation with an angle of plate current flow equal to 90°, that is, plate current flows for the full positive half cycle of the input voltage. The typical circuit diagram is the same as that for the Class C amplifier. Operation with respect to the $I_b - E_c$ characteristic is similar to that of Class C except that the grid is biased to practically plate current cutoff for any specified d-c plate potential. The efficiency of operation is about the same as a Class B audio (untuned) amplifier and is less than the efficiency of a Class C amplifier. However, as will be seen later, the special property of a Class B amplifier, when properly adjusted, is the nearly linear relation between the output voltage, or output current, and the input voltage. This property makes a Class B amplifier suitable for a power amplifier in a radio transmitter after amplitude modulation has been produced. An amplifier operating with more than cutoff bias causes modulation distortion because the output voltage is not linear with respect to the input voltage. In general, a Class C amplifier will increase the percentage of modulation and will cause over modulation for an input voltage that is nearly 100 per cent modulated because no plate current will flow when the instantaneous grid voltage is below cutoff voltage.

When a Class B amplifier is suitably adjusted for increasing the power level of a modulated source and an a-c voltage of constant amplitude is applied to the input the plate current is composed of nearly half sine wave pulses each of which occurs when the grid voltage is positive with respect to the bias voltage. A Fourier analysis for this kind of a current shows that the component of current of fundamental frequency has a peak amplitude equal to half the peak amplitude of the half sine wave pulses. The fundamental component of current flowing through the output impedance of the amplifier, which is tuned to the operating frequency, produces an a-c plate potential that is nearly a sine function. There are some harmonics but these are relatively small and do not greatly alter the analysis for the power output at the operating frequency. Under these conditions the total plate current is given by the expression

$$i_b = g_{gp}e_g + g_p e_p \text{ for } e_g + E_{cc} > 0$$

and $$i_b = 0 \text{ for } e_g + E_{cc} < 0$$

since $$i_{p1} = \tfrac{1}{2}i_b \text{ and } e_p = -i_{p1}R_{o1}$$

where i_{p1} is the a-c plate current component of the operating frequency and R_{o1} is the impedance of the tank circuit at the operating frequency.

From these relations $I_{p1} = \dfrac{\mu E_g}{2r_p + R_o}$ and $E_o = \dfrac{\mu E_g R_o}{2r_p + R_o}$
where E_o and E_g are either rms or peak values.

These expressions for the fundamental component of I_b and E_o can yield reasonable accuracy only when the dynamic curve of i_b and e_b is nearly a straight line.

In order that a Class B amplifier raise the power level of a modulated wave without distortion there must be a linear relation between the envelope of the output voltage and the envelope of the input voltage. This means the dynamic relation between I_b and E_c must be nearly a straight line over the complete range of E_c from cutoff to the positive peaks. Hence any graphical analysis for the voltage or power output must be carried out for several values of input voltage from zero to twice the carrier in order to make sure of the linearity of the amplifier to twice the carrier voltage. This involves a large amount of work even if one were certain at the start what region of the characteristic curves would give a nearly linear relation. Some of the work can be reduced by the following procedure. First determine the point of maximum i_b and minimum E_b on the E_c - E_b constant current chart for four times the desired carrier power output. This is carried out as described for Class C operation except for an angle of flow of 90°. Then draw a line from this point to the point where nearly plate current cutoff cuts the assumed d-c voltage E_{cd}. Next plot a curve of plate current versus grid potential, values to be taken from E_c - E_b chart. If this curve is nearly a straight line the assumed operation will nearly give linearity between E_o and E_g. On the other hand if this curve departs seriously from a straight line linearity between E_o and E_g cannot be expected. A new set of conditions must be assumed and the procedure repeated until a nearly linear curve between I_b and E_c is found. And finally having found a suitable region of operation the a-c power output can be determined for several constant values of a-c input voltage from zero to twice the voltage of the carrier. If the a-c power is proportional to the square of the input voltage, or nearly so, the amplifier will perform in a satisfactory manner. However there still remains the determination of the d-c power input and efficiency.

In the above graphical analysis a d-c operating voltage was assumed. If a preassigned efficiency of operation is assumed the relation between the d-c plate voltage and the peak value of the a-c plate voltage becomes fixed. Practical efficiency factors for carrier input voltage conditions range between 0.3 and 0.35. For an input voltage equal to twice the carrier voltage the efficiency factor is doubled. The relation between E_{bd} and E_{pc} for carrier condition is $E_{pc} = 1.27\ \mathcal{E}_c E_{bd}$, when there is linearity, where $\mathcal{E}_c$ is the efficiency factor. For twice carrier conditions $E_{p2c} = 2.54\ \mathcal{E}_c E_{bd}$. For a carrier efficiency factor of 1/3, $E_{p2c} = 0.84\ E_{bd}$, where E_p is the peak value of the a-c plate voltage for an a-c input voltage equal to twice the carrier voltage. This is the relation by which the d-c plate voltage can be located for the assumed peak a-c plate voltage or vice versa.

3-22. Neutralization of a Tuned Power Amplifier.- When a triode tube is operated with a parallel circuit of L, R, and C, in the plate circuit or any circuit that may be inductive at or near the frequency of any resonant tank circuit on the grid side of the tube, there is an excellent chance of self-oscillation unless proper methods are used to prevent them. It was pointed out in Sec. 3-3 that the grid-to-plate capacitance of a triode caused the input impedance to have a negative resistive component when the output impedance is inductive. The negative resistance means power is fed back into the grid circuit in such a way that the necessary power to drive the grid from an external source can be reduced or made equal to zero. When the external power needed is zero the tube becomes self-excited and ceases to be of any value as an amplifier.

The way to overcome this effect in a triode amplifier is to arrange a connection between the grid and plate that will result in a positive resistance component which will effectively cancel out the negative resistance component caused by the grid-to-plate capacitance. There are several ways of doing this. The simplest is to connect a coil in series with a blocking condenser between the plate and grid. The reactance of the coil is made equal to the reactance of the grid-to-plate capacitance at the operating frequency. The more successful methods are illustrated in Fig. 3-58. Either of these circuits is equivalent to a balanced bridge circuit. In either case a positive resistance is introduced into the grid circuit through the neutralizing condenser C_n which is just equal to the negative resistance caused by the grid-to-plate capacitance C_{gp}. Best

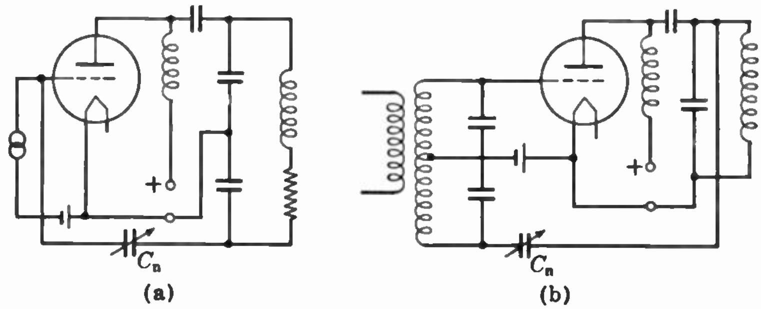


FIG. 3-58. Two methods for neutralizing the effect of the grid to plate capacitance in a triode tuned power amplifier.

results are obtained when the split tank condenser to which the neutralizing circuit is connected has equal capacitances on both sides of the neutral. Under these conditions $C_n = C_{gp}$. The system of Fig. 3-58b is more economical for large single-ended power tubes because it does not require a split tank condenser in the plate circuit. For push-pull amplifiers there would be no difference in cost because there is a neutral or ground connection either on the tank coil or tank condenser.

One method for adjusting the neutralizing condenser in practice is to apply exciting voltage to the grid with the plate voltage removed. Then with the aid of a resonance indicator such as a few turns of wire and milliammeter or wavemeter coupled to the tank coil tune the tank circuit to resonance; if necessary throw the neutralizing condenser off so as to get a good indication on the resonance indicator. Then finally adjust the neutralizing condenser until there is no indication of current in the tank coil and the circuit will be properly neutralized.

3-23. Frequency-Multiplier Amplifiers.- Most Class C power amplifiers are driven by other Class C amplifiers called buffer amplifiers and these buffer amplifiers in turn are excited from quartz crystal controlled oscillators. It is not practical to grind quartz crystals for stabilizing the frequency of an oscillator in excess of about 10 megacycles. Consequently when a higher stabilized frequency than this is wanted the crystal oscillator is adjusted for some sub-multiple of the desired frequency of the final power amplifier. Then one or more of the Class C amplifiers between the oscillator and final amplifier are operated as frequency doublers or triplers. That is, they are operated as frequency multiplier amplifiers where the output frequency is two or three times the input frequency.

In a frequency-multiplier amplifier the output tank circuit is adjusted to an integral multiple of the input frequency. For these conditions then the tank circuit offers a high impedance to a harmonic frequency of the input and very low impedances to the fundamental and all the other harmonic frequencies. The a-c plate voltage of the tube is nearly sinusoidal but a multiple frequency of the input voltage. For example in a frequency doubler the tank circuit is tuned to twice the frequency of the input, or exciting, voltage of the tube. Hence if the exciting voltage is $E_g \sin \omega t$ the plate voltage will be $E_p \sin 2\omega t$. Minimum plate voltage occurs when the grid voltage is a maximum because the instantaneous plate current is a maximum. The voltage and current relations are illustrated in Fig. 3-59.

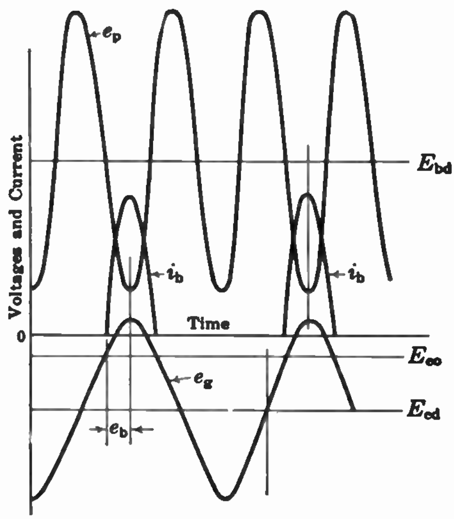


Fig. 3-59. Current and voltages in a frequency doubler.

In a frequency-multiplier amplifier the d-c operating and exciting voltages are adjusted so that plate current flows for 180° or less of the plate voltage cycle. If plate current is allowed to flow during the same fraction of the input voltage cycle as in the case of the normal Class C amplifier a part of the plate current will flow when the plate voltage is equal to or higher than the d-c voltages, with the result that the plate losses will be excessive and the efficiency low. Hence the angle of plate current flow with respect to the input cycle for a frequency multiplier is about equal to 90°/n, where n is the ratio of the output frequency to the input frequency. It is to be remembered that the angle θ_p of plate current flow is used for only half of the total time during which plate current flows. To bring about a smaller angle of plate current flow and still operate with about the same i_b maximum as in the case of the normal Class C amplifier requires a higher grid-bias voltage for the frequency-multiplier amplifier.

The power output of a frequency multiplier is about 1/n times that of a normal Class C amplifier because only one plate current pulse, which lasts for about 180° of the plate voltage cycle, flows for every n cycles of the plate voltage. The alternating component of plate current that has the same frequency as the plate voltage is about equal to 1/n of the corresponding current in a normal Class C amplifier. The d-c power input is also reduced a corresponding amount.

BIBLIOGRAPHY

"Radio Engineering," Terman, McGraw-Hill.
"Principles of Radio Engineering," Glasgow, McGraw-Hill.
"Communications Engineering," Everitt, McGraw-Hill.
"Radio Engineering Handbook," Henney, McGraw-Hill.
"Theory of Thermionic Vacuum Tubes," Chaffee, McGraw-Hill.
"Theory and Applications of Electron Tubes," Reich, McGraw-Hill.
"Fundamentals of Vacuum Tubes," Eastman, McGraw-Hill.
"High Frequency Alternating Currents," McIlwain and Brainerd, Wiley.
"Wide Band Television Amplifiers," Everest, Electronics, Jan. and May, 1938.
"Wide Band Amplifier for Television," Wheeler, Proc. I.R.E. July 1939.
"Simplified Methods for Computing Performance of Transmitting Tubes," Wagener, Proc. I.R.E. Jan. 1937.
"Stabilized Feed-back Amplifier," Black, Bell System Tech. Journal, Jan. 1934.
"Wide Band Amplifiers and Frequency Multiplication," D.L. Jaffee, Electronics, April 1942.

Chapter 4

TRIGGER CIRCUITS (GATES), PULSE-SHARPENING CIRCUITS AND OSCILLATORS

Trigger circuits, first described in 1919,[1] have become increasingly useful in recent years and now find many applications as the basis of electronic switches and pulse-forming circuits. Since all trigger circuits may also form the basis of relaxation and negative-resistance oscillators, a discussion of trigger circuits serves as a logical introduction to a treatment of these types of oscillators. Trigger circuits may be defined as circuits which, for fixed values of applied voltage and circuit parameters, have two stable conditions of equilibrium. The currents and voltages in such circuits can be made to change abruptly from one set of stable values to the other set of stable values at a critical value of some resistance or impressed voltage, and to change back to approximately their original values at a different critical value of resistance or impressed voltage.

4-1. Theory of Trigger Circuits.- The criterion as to whether a circuit element can serve as the basis of a trigger circuit can be determined from the characteristic current-voltage curve of the element. In Fig. 4-1, W represents the circuit element, R a resistance in series with the element, and E_b the supply voltage. The voltage across the element W is given by the equation

$$E = E_b - RI \qquad (4\text{-}1)$$

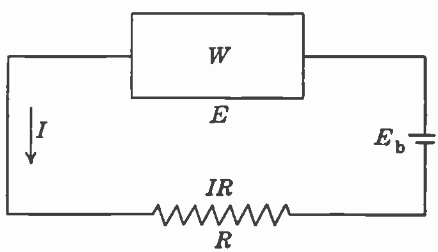


FIG. 4-1. Basic trigger circuit.

The voltage across the element is also a function of the current through the element, which may be indicated by

$$E = f(I) \qquad (4\text{-}2)$$

Since the form of eq. (4-2) is not in general simple, equilibrium values of current may be found most readily graphically, as shown in Fig. 4-2. Equation (4-2) represents the characteristic curve of the element W. Equation (4-1) is that of a straight

1. Eccles, W.H., and Jordan, F.W., Radio Rev., 1, 143 (1919).

line MN through a point on the voltage axis corresponding to the supply voltage E_b, having a negative slope in amperes per volt equal to the reciprocal of the resistance in series with the element. Equilibrium values of current are determined by the intersection of the characteristic curve with the resistance line.

It can be seen from Fig. 4-2 that, if the characteristic curve has a portion whose slope is negative, the resistance line may intersect the curve in three points: 1, 2, and 3, indicating that there are three possible equilibrium values of current. Since, with E_b and R constant, an increase of current from the value represented by point 2 would be accompanied by a decrease of voltage across the element, more voltage would thus be made available to send current through the resistance, and the current would rise further. Conversely, any decrease in current through the element would reduce the voltage available across the resistance and thus cause a further reduction of current. Point 2 therefore lies in a range of unstable equilibrium and is not observed experimentally. If the applied voltage is raised progressively from zero, the intersection moves along the branch OA of the characteristic curve. When the intersection is at A, an infinitesimal increase of voltage causes the current to fall abruptly to the value at E. Further increase of supply voltage causes the intersection to rise toward C. If the battery voltage is then decreased continuously, the intersection moves down the branch CB until the point B is reached, at which the current jumps abruptly to the value corresponding to point D. It can be seen that similar abrupt changes of current result if the slope of the resistance line is varied by changing the resistance R, or if the characteristic curve is displaced vertically or horizontally. With trigger circuits incorporating vacuum tubes, this displacement can be accomplished by varying one or more electrode voltages.

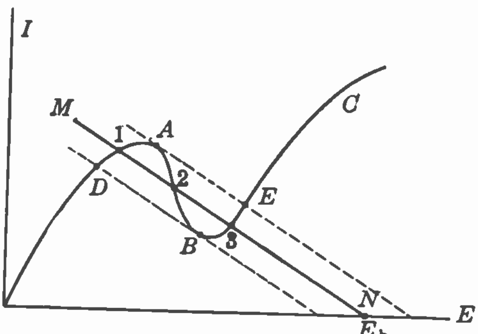


FIG. 4-2. Negative-resistance characteristic.

From the above analysis it follows that a circuit element whose current-voltage characteristic has a portion with negative slope may serve as the basis of a trigger circuit. This is equivalent to saying that the element must have a negative a-c resistance over a portion of its current range.

4-2. Tetrode Trigger Circuits.- The plate characteristics of screen-grid tetrodes are similar in form to the curve of Fig. 4-2. Hence it is to be expected that a trigger circuit can be formed by introducing a resistance in series with the plate supply voltage. The slope of the negative-resistance portion of the plate characteristics of modern tetrodes is so low, however, that series resistances of the order of 100,000 ohms or greater, and correspondingly high values of plate supply voltage must be used. For this reason, and because of change in shape of the characteristic with tube age as the result of changes in secondary emission, this type of circuit has found little or no application. The student should, however, keep it in mind as a possible tool in future problems. The circuit can be triggered by varying any of the electrode voltages or the series resistance.

4-3. Pentode Trigger Circuits.— Figure 4-3 shows the characteristic relating the screen current i_{c2} of a pentode with the screen voltage e_{c2} when the suppressor is connected to the screen in such a manner that a change in screen voltage is accompanied by a proportional change of suppressor voltage, the suppressor voltage being maintained negative and the plate positive.[2] Since this curve is of the general form of that of Fig. 4-2, the introduction of resistance in series with the screen supply voltage of a pentode with screen-suppressor coupling results in a trigger circuit. Figure 4-4 shows such a circuit, in which resistance coupling is used between the screen and the suppressor.[3] The circuit may be triggered by means of voltages in series with any of the electrode supply voltages, by voltage pulses impressed upon one of the grids through a condenser, as shown in Fig. 4-4, or by changes of circuit resistances. The control grid is the most sensitive electrode for the purpose of triggering. The values of the supply voltages are not critical, but the proper relation must be maintained between them. The circuit constants shown in Fig. 4-4 are typical.

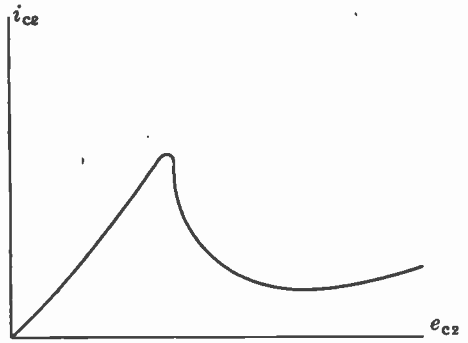


FIG. 4-3. Characteristic of screen current i_{c2} vs screen voltage e_{c2} for pentode, change Δe_{c3} in suppressor voltage $\propto \Delta e_{c2}$.

2. Herold, E.W., Proc. I.R.E., 23 1201 (1935).
3. Reich, H.J., Rev. Sci. Instr., 9, 222 (1938).

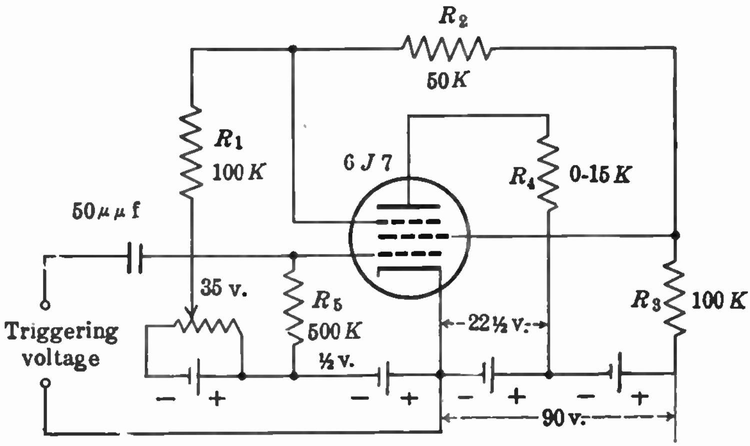


FIG. 4-4. Pentode trigger circuit.

A physical explanation of the operation of the circuit is not difficult. Suppose that the screen current has its lower equilibrium value, and that the screen voltage is reduced. The reduction of screen voltage tends to reduce the screen current, but, because of the coupling to the suppressor, it also makes the suppressor more negative. The increase of negative suppressor voltage decreases the plate current and thus increases the screen current, since electrons repelled by the suppressor are returned to the screen. In certain ranges of operating voltages the increase in screen current resulting from the increase of negative suppressor voltage is greater than the decrease in screen current resulting from the decrease in screen voltage, and so the net result is an increase of screen current. Since increased screen current is accompanied by a further drop in screen voltage as the result of RI drop in the screen circuit resistance, the action may become cumulative and the screen current rise to its higher stable value. The process may then be reversed by increasing the screen voltage. Because an increase of screen current is accompanied by a reduction of plate current, the plate current also has two stable values, the higher of which corresponds to the lower value of screen current. By proper choice of voltages and circuit constants, the lower value of plate current may be made zero. The possibility of making the lower value of plate current zero is important in certain applications of the circuit.

4-4. Eccles-Jordan Trigger Circuit.- The best-known and most useful trigger circuit is that devised by Eccles and Jordan and shown in basic form in Fig. 4-5. This circuit functions by virtue of the fact that only one tube at a time passes plate current. This may be readily shown by a physical analysis of the circuit. Let it be assumed that an equilibrium condition exists in which both tubes conduct simultaneously. Then an increase of current in either tube increases the voltage drop in the corresponding coupling resistance and thus increases the negative grid voltage of the other tube and reduces the plate current of the other tube. This in turn reduces the negative grid voltage of the first tube and causes further increase of plate current of the first tube. The action is cumulative, and so the current of one tube falls to zero.

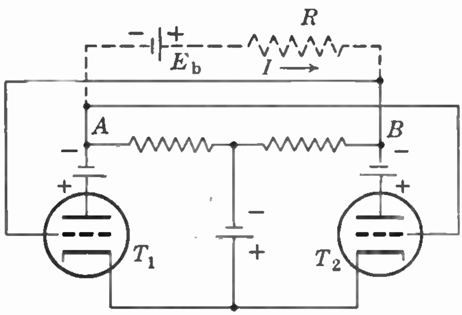


FIG. 4-5. Basic Eccles-Jordan trigger circuit

Triggering of the Eccles-Jordan circuit is also predicted by an analysis based upon a current-voltage characteristic of the circuit. Figure 4-6 shows a curve of external current that flows as the result of the application of direct voltage between points A and B of the circuit of Fig. 4-5. Over a limited range of impressed voltage the current flows in the direction opposite to that in which the applied voltage alone would cause it to flow. Between points A and B, therefore, the circuit acts like a negative resistance. When a battery E_b is connected to the points A and B through a resistance R, as shown by the dotted lines of Fig. 4-5, the corresponding resistance line is the line MN in the current-voltage diagram of Fig. 4-6. If R exceeds in magnitude the value of the reciprocal of the slope of the curve at point O, abrupt changes of current through R and of voltage

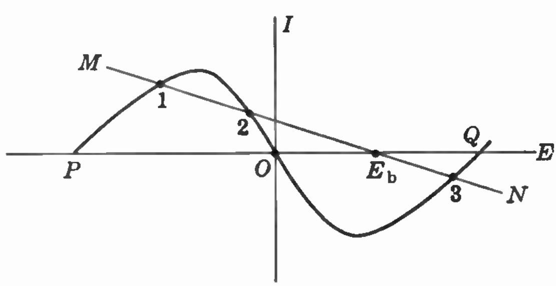


FIG. 4-6. Current-voltage charateristic between points *A* and *B* of the circuit of Fig. 4-5.

between A and B can be made to occur by varying E_b or by shifting the characteristic curve by changing the operating voltages of the tubes. If R is increased, MN becomes more nearly horizontal and, in the limiting case, when R is infinite, coincides with the voltage axis. The external current is then zero, but changes in electrode voltages can cause an abrupt transfer of current from one tube to the other and a reversal of voltage between A and B, corresponding to the intercepts P and Q of the characteristic on the voltage axis. The equilibrium voltage between A and B is then equal to the product of R_b and the equilibrium plate current of one tube. Ordinarily this circuit is used without the external resistance R and the battery E_b, the circuit being triggered by changes of grid or plate voltages of such polarities as to reduce the plate current of the conducting tube or increase the plate current of the non-conducting tube. It should be noted, however, that the use of a voltage E_b and of an external resistance R only slightly lower than the magnitude of the reciprocal of the slope of the curve at O makes possible an abrupt reversal of current through the resistance as the result of only a very small change in the voltage E_b. This method of using the circuit is sometimes advantageous.

The need for more than one voltage supply is avoided in the circuit of Fig. 4-7, in which the coupling between tubes is made by means of resistances of proper values to maintain the correct operating grid voltages. This circuit may be triggered by means of voltages introduced in series with the electrodes, by changes of the circuit resistances, or by means of voltage pulses applied to one or more electrodes through a transformer or through a condenser as shown in Fig. 4-7.

The grid-cathode capacitance of the tubes in the circuit of Fig. 4-7 tends to prevent the grid voltage from changing

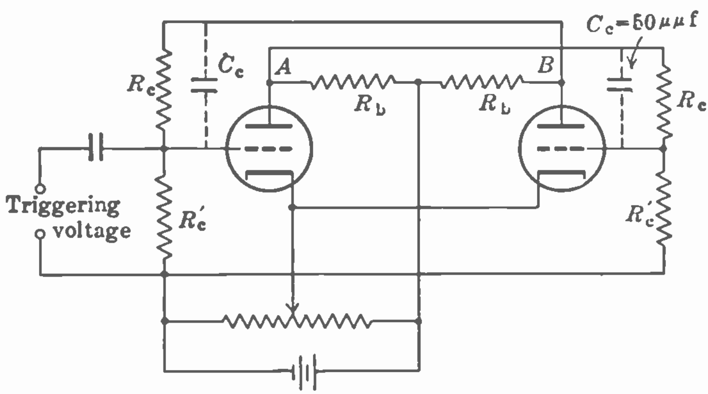


FIG. 4-7. Practical form of Eccles-Jordan trigger circuit.

relative to the cathode. The grid-plate capacitance couples the grids to the plates of the same tubes and thus tends to cause the grid voltage to change in the direction to prevent triggering. The plate-cathode capacitance tends to prevent the plate voltage from changing relative to the cathode. The interelectrode capacitances therefore act in such a manner as to increase the triggering time and to reduce the reliability of operation. The effects of the interelectrode capacitances may be offset by the coupling condensers, C_c, the capacitance of which is of the order of 50 μμf. Because the voltage across these condensers cannot change instantaneously, a change of voltage of either plate results in a nearly equal instantaneous change of voltage of the grid of the other tube. Without the coupling condensers the voltage dividing action of the coupling and biasing resistors R_c and R_c' prevents the change in grid voltage from exceeding about one-half the change of plate voltage, and the action of the interelectrode capacitances reduces the change in grid voltage considerably below this value.

Another modification[4] of the basic Eccles-Jordan circuit is shown in Fig. 4-8, in which the suppressor grids of pentodes serve the same function as the triode control grids of the circuit of Fig. 4-7. A constant voltage is applied to the screen grids, as in the use of pentodes as voltage amplifiers. The

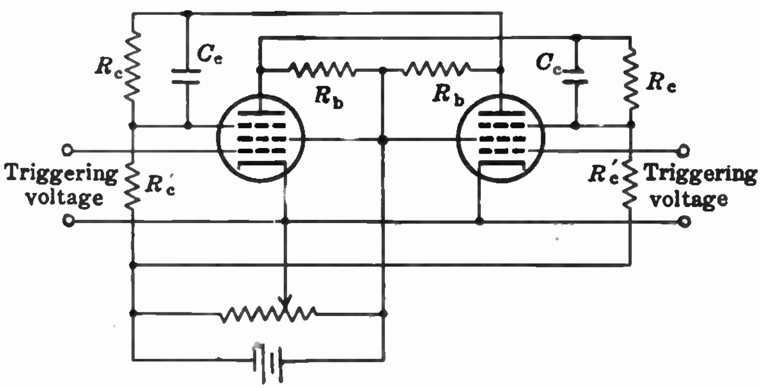


FIG. 4-8. Modified Eccles-Jordan trigger circuit using pentodes.

control grids are used for triggering the circuit. Because the high negative voltage of the suppressor of the non-conducting tube prevents the flow of plate current regardless of the voltage of the control grid, the circuit cannot be made to trigger

4. H.J. Reich, loc. cit.

by the application of positive voltage to the control grid of the non-conducting tube. The application of a negative pulse of voltage to the control grid of the conducting tube, however, reduces its current to zero and thus triggers the circuit. If short negative voltage pulses are applied simultaneously to both grids, the plate currents of both tubes remain zero during the duration of a pulse. By applying a higher negative voltage to the suppressor of the tube which has been conducting, the condensers C_c insure the transfer of current to the other tube at the end of the triggering pulse. The size of the condensers should be such that the time taken for them to charge or discharge from one equilibrium value of voltage to the other value is large in comparison with the duration of the triggering impulse, but small in comparison with the time between successive pulses. Condensers of 50 μμf capacitance are usually satisfactory. Reliable triggering necessitates the use of triggering pulses of very short duration.

The functions of the control and suppressor grids in the circuit of Fig. 4-8 may be interchanged, but the circuit is then sensitive to triggering voltage of either polarity. It is also possible to connect the control and screen grids of pentodes as the triode grids and plates respectively are connected in the circuit of Fig. 4-7, and to take the output voltage or current from the plate circuits.

The values of the resistances R_b, R_c, and R_c' are not critical in the circuits of Figs. 4-7 and 4-8, but R_c and R_c' should be appreciably larger than R_b, and may usually be equal to one another. The plate currents decrease with increase of R_b, and the voltage across R_b increases. The choice of R_b is therefore governed to some extent by whether a current or voltage output is desired from the circuit. Typical resistance values are R_b = 10,000 ohms and $R_c = R_c'$ = 250,000 ohms. Supply voltages as low as 22½ volts may be used in these circuits, but the voltage output available across R_b increases with supply voltage.

In using trigger circuits, any of the circuit currents may be used directly to operate a relay or other current-controlled device, or the voltage drop across one of the resistors may be used directly or applied to the grid of another tube to control its plate current. The circuits may be triggered by means of voltages applied to one or more electrodes, as already explained. Usually the circuits are so sensitive that the change of electrode voltage resulting from touching one of the tube electrode terminals is sufficient to cause triggering. The circuit of Fig. 4-8 can be triggered by negative control grid voltage as low as ½ volt. The Eccles-Jordan circuits may also be triggered by means of changes of illumination if phototubes are connected in series with or in place of the resistances R_c or in parallel with or in place of resistors R_c'.

4-5. Circuits for Generating Rectangular Voltage Pulses.- Rectangular waves of voltage are produced across the screen or plate resistors of the circuit of Fig. 4-4 and across the plate resistors of the circuits of Figs. 4-7 and 4-8 when the circuits are triggered periodically. If the successive pulses are spaced so that the circuits remain in one equilibrium state longer than in the other, the positive and negative halves of the rectangular waves are of unequal duration. Random rectangular pulses may be generated by the application of two triggering impulses separated in time by the required duration of the rectangular pulse. The first impulse triggers the circuit in one direction, and the second in the other.

Periodic or random rectangular voltage or current pulses of controllable length initiated by single triggering impulses may also be generated by unbalancing a trigger circuit in such a manner that it normally remains in one equilibrium state, and returns to that state after a short time interval if triggered into the other equilibrium state. This may be accomplished in the Eccles-Jordan type of circuit by the use of unequal bias on the grids of the two tubes. The grid of one tube may have zero bias or a small positive bias, the flow of high grid current being prevented by the grid coupling resistor. In order that the circuit return to its normal state after triggering, one or both of the coupling resistors R_c are eliminated, and a comparatively large value of coupling capacitance C_c is used. A typical circuit is shown in Fig. 4-9.

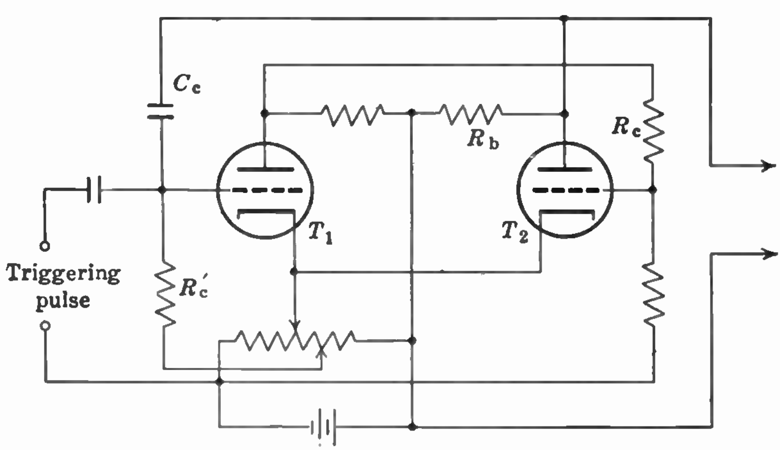


FIG. 4-9. Circuit for generating rectangular voltage pulses.

The operation of the circuit of Fig. 4-9 is as follows. The voltage divider controlling the grid bias of tube T_1 is adjusted so that this tube normally carries plate current, but so that the circuit can be triggered. The application of a negative triggering impulse to the grid of T_1 or of a positive impulse to the grid of T_2 stops the plate current in T_1 and starts it in T_2. The resulting RI drop in R_b lowers the plate voltage of T_2. Since the voltage of the condenser C_c cannot change instantaneously, the grid voltage of T_1 is lowered by an amount that is initially equal to the reduction of plate voltage of T_2, and current continues to flow in T_2. The drop in voltage across R_b, however, causes current to flow into C_c through R_c' and thus raises the grid voltage of T_1. After a time interval that is determined principally by the magnitude of the time constant $R_c'C_c$, and to a lesser extent by the value of R_b and the supply voltage, the grid voltage of T_1 reaches a value at which the circuit again triggers, and current stops flowing in T_2. If the resistance R_b is sufficiently small in comparison with R_c' so that the condenser current does not greatly affect the RI drop in R_b, the voltage pulse produced across R_b is rectangular in form. The length of the pulse is controlled by varying $R_c'C_c$. Obviously, periodic pulses of desired frequency may be produced by the use of periodic triggering pulses. Similar results are obtained if the second coupling resistor, R_c, is replaced by a condenser, but there is then the possibility that the circuit will oscillate in a manner to be explained in Sec. 4-9.

The use of rectangular pulses ("square waves") in circuit testing is discussed in Ch. 1.

4-6. **Circuits for Generating Triggering Pulses.**- In order to insure reliability in triggering, the triggering impulses applied to trigger circuits must be pulses of very short duration. Such a pulse may be produced by the simple RC circuit of Fig. 4-10 if R and C are small. If the direct voltage impressed upon the input to the circuit is changed abruptly, the condenser charges or discharges exponentially until its voltage equals the new value of impressed voltage. Voltage appears across the resistance only during the charging time. If the time constant RC is small, the charging time is short and an exponential pulse of short duration appears across the output of the circuit. A typical pulse[5] is shown in Fig. 4-11. If the input voltage is a

5. The form of the pulse, assuming all inductance in the circuit may be neglected, may be derived from the differential equation of the simple RC circuit and is $Ri = Ee^{-t/RC}$. The effect of a sudden change in applied voltage on more complex linear circuits usually requires analysis by a Fourier Integral or similar method, as outlined in Ch. 1. In the case of very short pulses it may be extremely difficult to obtain the sharp rise shown in Fig. 4-11.

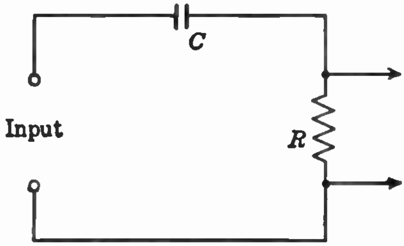


FIG. 4-10. Circuit for generating triggering pulses (differentiating circuit).

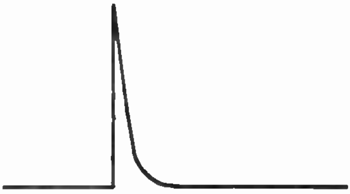


FIG. 4-11. Form of pulse produced by an abrupt change of direct voltage in the input of the circuit of Fig. 4-10.

periodic wave having discontinuities, such as a rectangular or triangular wave, an output pulse is produced at each discontinuity, as shown in Fig. 4-12.

Sharp pulses may be derived from a sinusoidal voltage by clipping off the peaks of the wave by means of some form of voltage limiter before impressing it upon the circuit of Fig. 4-10. Figure 4-13a shows a simple biased diode circuit by means of which a sine wave may be converted into a wave approximating rectangular form. The output voltage increases with input voltage until the instantaneous impressed voltage is approximately equal to the biasing voltage. Current then starts flowing in one of the diodes and the RI drop in the series resistor prevents appreciable rise of output voltage. The solid curve of Fig. 4-13b shows the form of the output voltage. The output voltage may be amplified and applied to a similar diode circuit in order to make the changes in voltage more abrupt. By the use of a number of limiter stages the

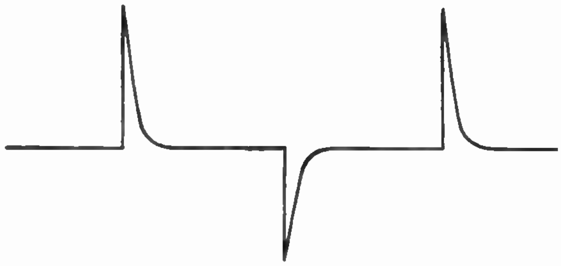


FIG. 4-12. Output voltage of the circuit of Fig. 4-10 for square-wave input.

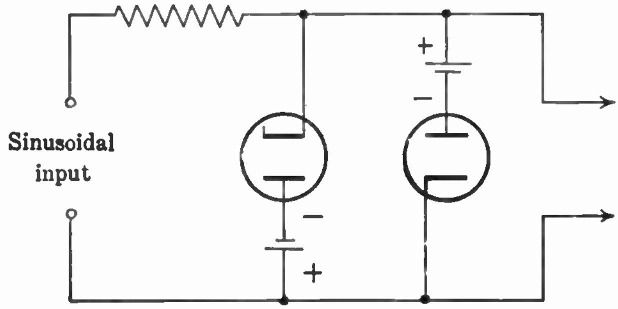


FIG. 4-13*a*. Peak-clipping circuit.

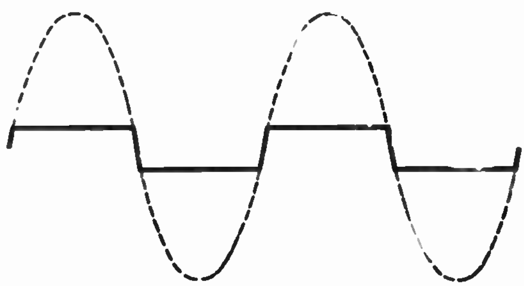


FIG. 4-13*b*. Output voltage of the circuit of Fig. 4-13*a*.

voltage may be made to approach as nearly as desired a true rectangular form. Application of this voltage to the pulse-sharpening circuit of Fig. 4-10 converts the square wave into periodic pulses which are alternately positive and negative and may be used for triggering. Peak clipping may also be accomplished by the flow of grid current in an amplifier tube.

A symmetrical voltage pulse may be obtained by use of the circuit of Fig. 4-14, which is a modification of that of Fig. 4-13. In this circuit the bias is made so high that current flows through the diodes only at the peaks of the voltage waves, and the output voltage across the resistance R is used for output voltage. If only one diode is used, the pulses are of one polarity only; if two diodes are used, as in Fig. 4-14, alternate pulses are of opposite polarity. By amplifying the output voltage and applying it to a similar circuit, the sharpness of the pulses can be increased. The duration of the pulses can be decreased to any desired point by the use of a number of stages.

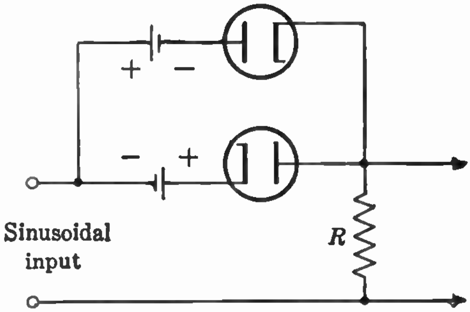


FIG. 4-14. Circuit for generating symmetrical periodic pulses.

Sharp periodic pulses of voltage can also be generated by a self-biased sinusoidal oscillator in which the biasing resistor and condenser are so large that oscillation ceases after a single cycle of oscillation and does not start until the lapse of an appreciable time. The action of such a circuit is explained in Sec. 4-24. Figure 4-15 shows a typical circuit, and Fig. 4-16 a typical wave of voltage generated by the circuit. If the resonance frequency of oscillation of the circuit is high, the pulse is of very short duration. The repetition frequency is governed by the product RC.

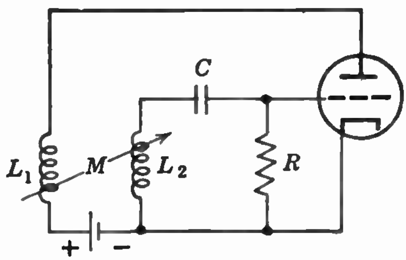


FIG. 4-15. Regenerative pulse-generating circuit. (blocking oscillator).

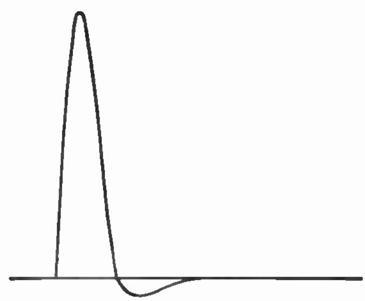


FIG. 4-16. Form of pulse generated by the circuit of Fig. 4-15.

4-7. Relaxation Oscillators.- Relaxation oscillators are oscillators in which one or more currents or voltages change abruptly at one or more times in the cycle of oscillation. Any trigger circuit can be transformed into a relaxation oscillator by incorporating condensers in the circuit in such a manner that triggering is followed by the charging or discharging of one or more condensers. At a critical value of condenser voltage the circuit triggers back again. Relaxation oscillators are useful in many applications because of their high harmonic content, the ease with which they can be synchronized to some other source of alternating voltage by the introduction into the oscillator circuits of small amounts of the external voltage, the wide range of frequency that can be obtained in a single oscillator, and their compactness, simplicity, and low cost. They are also of great value in the production of saw-tooth voltages required for the operation of cathode-ray oscillographs. The high harmonic content of relaxation oscillators, on the other hand, makes them unsuitable for applications in which sinusoidal wave form is essential.

4-8. **Van der Pol Relaxation Oscillator.**- Figure 4-17 shows the circuit of the Van der Pol relaxation oscillator, which is derived from the pentode trigger circuit of Fig. 4-4 by replacing the screen-suppressor coupling resistor by a condenser. The action of the circuit is as follows. Triggering of the circuit from the higher to the lower value of screen current causes

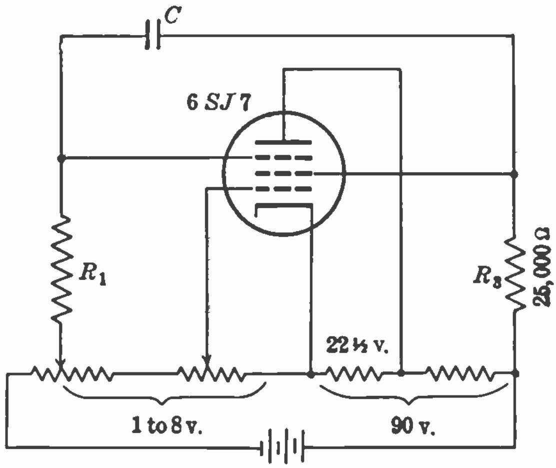


FIG. 4-17. Van der Pol oscillator.

an abrupt rise in screen voltage as the result of decreased RI drop in the screen resistor. Because the condenser voltage cannot change instantaneously, there is an initial change of suppressor voltage equal to the change in screen voltage. The decreased negative suppressor voltage maintains the lower screen current, but a charging current immediately starts flowing into the condenser through R_1 and the combination of R_3 in parallel with the screen-cathode path. As the voltage across the condenser rises, the suppressor voltage becomes more negative and at a critical value allows the circuit to trigger back to the higher value of screen current. The screen and suppressor voltages then fall abruptly, but as the condenser discharges, the suppressor voltage rises until the circuit again triggers. Figure 4-18 shows the wave forms of the condenser and suppressor voltages. The screen voltage is similar in form to the suppressor voltage. The wave is asymmetrical because during the charging of the condenser the screen current has its higher value and, if the suppressor swings positive, suppressor current may flow.

FIG. 4-18. Typical wave form of condenser current or suppressor voltage for the circuit of Fig. 4-17.

The flow of suppressor current and the increase of screen current cause the condenser voltage to change more rapidly than during the discharge of the condenser, when no suppressor current flows and the screen current is lower. The frequency of oscillation increases with decrease of R_1 and C_c and is also dependent upon R_3.

4-9. **Multivibrator.**- Figure 4-19 shows the circuit of the multivibrator, which is derived from the Eccles-Jordan trigger circuit of Fig. 4-7 by eliminating the coupling resistors R_c and increasing the capacitance of the coupling condensers C_c.

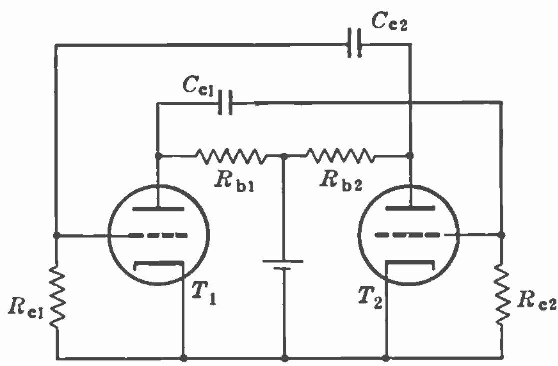


FIG. 4-19. Multivibrator.

Triggering of the circuit is followed by the flow of currents into the condensers through the resistors R_c' and R_b. The resulting change in condenser voltages reduces the grid bias of the non-conducting tube and raises the bias of the conducting tube, and finally causes the circuit to trigger back. If the circuit is completely symmetrical, the wave forms of the currents and voltages are symmetrical. Figure 4-20a shows a typical wave of grid and plate voltages and of condenser current for a symmetrical circuit. Figure 4-20b shows a typical wave of condenser voltage. The period of one-half of the cycle varies with the time constant $(R_{c1}' + R_{b2})C_{c2}$ and that of the other half of the cycle with the time constant $(R_{c2}' + R_{b1})C_{c1}$. If one time constant is made much smaller than the other, one-half of the cycle will be correspondingly shorter than the other.

FIG. 4-20a. Wave form of condenser current of symmetrical multivibrator.

FIG. 4-20b. Wave form of condenser voltage of aymmetrical multivibrator.

4-10. Synchronization of Relaxation Oscillators.- Introduction into relaxation oscillator circuits of small voltages having the same frequency, a multiple, or submultiple of the oscillation frequency synchronizes the relaxation oscillator to the control frequency. Small changes of oscillator circuit constants or voltages then do not change the frequency of oscillation. With reasonable care in circuit adjustment, relaxation oscillators may be controlled when the frequency ratio is as great as 50, but in order to insure reliability of operation, it is best not to exceed a ratio of about 10. The usual method of introducing synchronizing voltage into relaxation oscillator circuits is by means of transformer or condenser coupling to one or more of the grid circuits.

4-11. Saw-tooth Wave Generators.- By proper adjustment of circuit voltages and constants the condenser voltage in the circuits of Figs. 4-17 and 4-19 may be made to assume one of the forms illustrated in Fig. 4-21. Figure 4-22 shows another circuit, based upon the pentode trigger circuit of Fig. 4-4, with which a wave of the form of Fig. 4-21b can be more readily obtained. The output voltage is taken from across the condenser C, and the frequency is varied by means of R and C. Saw-tooth voltages having frequencies up to about 50,000 cycles per second are usually obtained by means of arc-tube relaxation oscillators.

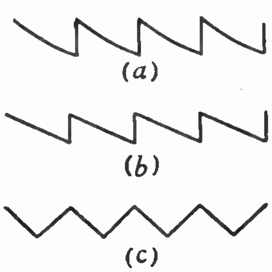


FIG. 4-21. Typical waves of condenser voltage in the circuits of Figs. 4-17 and 4-19.

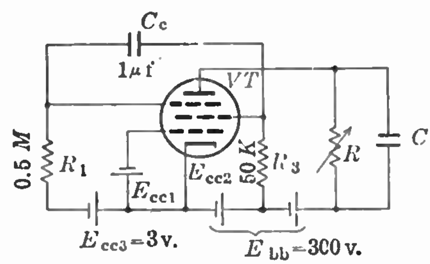


FIG. 4-22. Relaxation oscillator for the production of saw-tooth voltage wave.

4-12. Arc-tube Relaxation Oscillator.- Below a critical anode voltage, which is determined by the magnitude of the negative grid voltage, the anode current of a grid-controlled arc rectifier (thyratron or gas-discharge tube) is negligible. When the anode voltage reaches the critical value, however, the tube "breaks down" or "fires." The anode current is then determined only by the applied anode circuit voltage and the anode circuit impedance, and is independent of the grid voltage. The anode current can be interrupted only by reducing the anode voltage below the ionization voltage of the gas or vapor and keeping it there for a sufficiently long time to allow de-ionization to take place. In the operating range of current the anode voltage (tube drop), E_a, remains constant at about 10 volts in mercury-vapor tubes and about 16 volts in gas-filled tubes after the tube fires. The tube therefore behaves like a switch in series with a negligible impedance and a constant emf equal to the anode voltage.

Figure 4-23 shows the basic circuit of an arc-tube relaxation oscillator. The action of the circuit is as follows. The condenser C is charged exponentially through the high resistance R, the resistance of which is ordinarily of the order

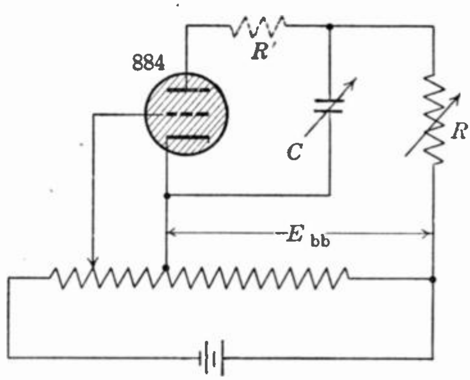


FIG. 4-23. Arc-tube relaxation oscillator.

of megohms. When the condenser voltage becomes equal to the firing voltage V_f of the tube, the tube starts conducting, and the condenser discharges through the tube. Because the current is limited only by the small circuit impedance in series with the condenser, the discharge time is very small in comparison with the charging time of the condenser. Ordinarily the ratio of the circuit inductance to the effective resistance is so small that the condenser discharges exponentially and the tube goes out when the condenser voltage has fallen to a value equal to the tube drop E_a, which is approximately equal to the ionization voltage of the gas or vapor contained in the tube. The condenser then charges again to the firing voltage and the cycle repeats. Figure 4-24 shows the manner in which the condenser voltage varies. Under the assumption that the discharging time of the

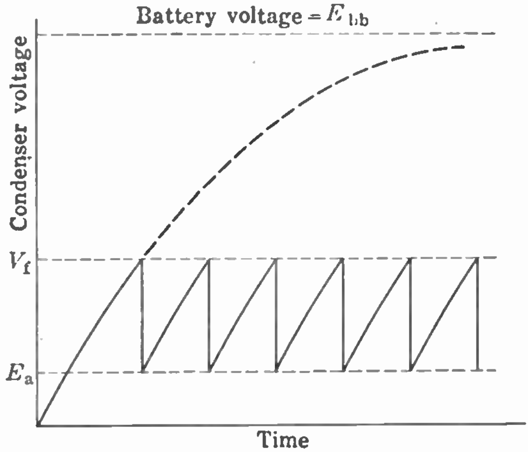


FIG. 4-24. Wave form of condenser voltage in the circuit of Fig. 4-23.

condenser is negligible in comparison with the charging time, the frequency of oscillation is

$$f = \frac{1}{RC \log_e \dfrac{E_{bb} - E_a}{E_{bb} - V_f}} \tag{4-3}$$

The frequency is varied by changing R or C. Large changes of frequency are usually made by changing condenser size and small changes by varying the resistance. Increasing the grid bias of the arc tube both reduces the frequency and increases the voltage amplitude. In order to prevent damage to the discharge tube as the result of excessive anode current during condenser discharge, or to change the wave form of the output voltage (see Sec. 5-8), a resistance R' is usually used between the condenser and the anode or cathode of the tube. For an 884 or 885 tube, this should have a minimum value of 1000 ohms.

If the supply voltage E_{bb} does not greatly exceed the breakdown voltage V_f of the tube, the wave of condenser voltage has appreciable curvature. If E_{bb} is large in comparison with V_f, on the other hand, the condenser voltage varies nearly linearly with time between the values E_a and V_f, and so the voltage produced across the condenser is of essentially saw-tooth wave form. The charging of the condenser can be made truly linear by maintaining the charging current constant in spite of change of condenser voltage. This can be accomplished by replacing the resistance R of the circuit of Fig. 4-23 by a voltage pentode, as shown in Fig. 4-25. At low values of anode current the current is independent of anode voltage for anode voltages exceeding about 20 volts, and so the charging current is constant. Under

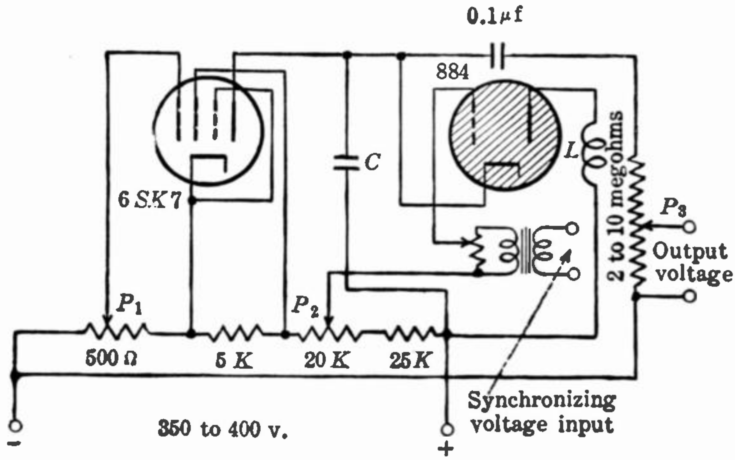


FIG. 4-25. Thyratron relaxation oscillator in which the condenser charging current is controlled by a pentode.

the assumption that the discharge time is negligible, the frequency of oscillation is

$$f = I/(V_f - E_a)C \qquad (4\text{-}4)$$

The frequency can be varied continuously in the circuit of Fig. 4-25 by means of the pentode control-grid bias.

The resistance R' in the circuit of Fig. 4-23 may be replaced by a small inductance L, as shown in Fig. 4-25. If the inductance in henrys is equal to the capacitance in microfarads, the maximum discharge current does not exceed the maximum discharge current that flows when a 1000-ohm resistance is used. When inductance is used, the tube does not go out when the condenser voltage becomes equal to E_a but, because of voltage induced in the inductance, continues to pass current until the condenser voltage has passed through zero and reversed. If the Q of the circuit is high, the magnitude of the maximum reverse voltage of the condenser is approximately $V_f - E_a$. The variation of condenser voltage is therefore nearly doubled when V_f is large. With a type 884 tube the maximum value of firing voltage is 300 volts (when the grid voltage is -30) and the condenser voltage varies between approximately -285 volts and +300 volts. The frequency of oscillation has the following value when inductance is used to limit the discharge current and the condenser charging current is controlled by a resistance R:

$$f = \frac{1}{RC \log_e \frac{E_{bb} + V_f - E_a}{E_{bb} - V_f}} \qquad (4\text{-}5)$$

When inductance is used and the charging current is maintained constant by means of a pentode, the frequency of oscillation is:

$$f = I/(2V_f - E_a)C \qquad (4\text{-}6)$$

Very low frequencies of oscillation can be attained in arc-tube oscillators by the use of large low-leakage condensers. The high-frequency limit at which an arc-tube saw-tooth wave generator is useful is limited in part by the time taken for the condenser to discharge. At frequencies above 10,000 cycles this time may be an appreciable part of the cycle. At very high frequencies the time taken for the tube to de-ionize also distorts the voltage from saw-tooth form. If residual ionization allows some anode current to flow at the beginning of the charging period, some of the current that should flow into the condenser flows through the tube, and so the condenser voltage does not rise as rapidly as it should. The distortion is less when inductance is used in series with the tube, since the condenser voltage, and hence the anode voltage, is negative at the time the tube is extinguished. The high negative anode voltage causes rapid de-ionization.

The output voltage from an arc-tube relaxation oscillator is usually taken from across the condenser. If a variable output voltage is required, the condenser may be shunted by a voltage divider, as shown in Fig. 4-25. In order to prevent diversion of appreciable current from the condenser and consequent loss of linearity of charging, the resistance of the voltage divider must be high. In order to increase the available output voltage and to prevent change of wave form and frequency as the result of current drawn by the load, the oscillator is usually followed by a direct-coupled or resistance-capacitance-coupled amplifier, which furnishes the required voltage, current, or power output. Voltage pulses of short duration may be obtained from across the current-limiting resistance or inductance used in series with the arc-discharge tube.

4-13. Synchronization of Arc-tube Oscillators.- Arc-tube oscillators, like high-vacuum-tube relaxation oscillators, may be synchronized to a control voltage whose frequency is approximately equal to a multiple or submultiple of the natural frequency of oscillation of the circuit. Synchronizing voltage is usually impressed in the grid circuit of the arc-discharge tube through a transformer, as shown in Fig. 4-25. If the natural frequency is slightly lower than the frequency of the control voltage, a positive peak of control voltage reduces the negative grid voltage enough to allow the tube to fire just before it would fire without control voltage. Since this happens during each cycle of the control voltage, the oscillator frequency is increased to that of the control voltage. If the oscillator frequency is the n'th multiple of the control frequency, the tube is caused to fire by the control voltage each n'th cycle of the oscillator. If the control frequency is the n'th multiple of the oscillator frequency, on the other hand, each n'th peak of the control voltage causes the tube to fire. In order to prevent possible distortion of the output voltage, only sufficient synchronizing voltage should be used to prevent the frequency from drifting.

4-14. High-frequency Saw-tooth-wave Generator.- Figure 4-26 shows a circuit that produces a saw-tooth wave at frequencies higher than those at which the arc-tube relaxation oscillator can function satisfactorily. In this circuit the triode is

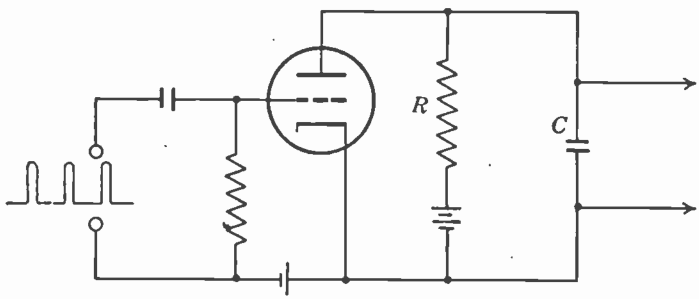


FIG. 4-26. High-frequency saw-tooth-wave generator.

normally biased beyond cutoff, but the application of periodic positive pulses to the grid causes it to become conducting throughout the duration of the pulses. The condenser charges through the resistor R and discharges rapidly through the tube each time a voltage pulse is impressed on the grid. The rate at which the voltage rises and the maximum voltage to which the condenser charges at a given frequency is controlled by the sizes of the condenser and the series resistance and by the battery voltage. The frequency is determined by the frequency of the pulses applied to the grid. In order to insure that the condenser voltage rises nearly linearly the supply voltage must be large in comparison with the maximum voltage to which the condenser charges. True linear rise of condenser voltage can be obtained by replacing the resistance R by a pentode, as in the arc-tube relaxation oscillator. In order to insure that the condenser starts discharging abruptly, the pulses impressed upon the grid of the triode should have a steep wave front. They may be generated by circuits of the types discussed in Secs. 4-5 and 4-6.

4-15. Circuit for Generating Periodic Triangular Pulses of Short Duration.- In certain applications of cathode-ray tubes it is necessary to make use of periodic triangular pulses of voltage of short duration separated by relatively long time intervals. This can be accomplished by the circuit[6] of Fig. 4-27, which is similar to that of Fig. 4-26. In this circuit the

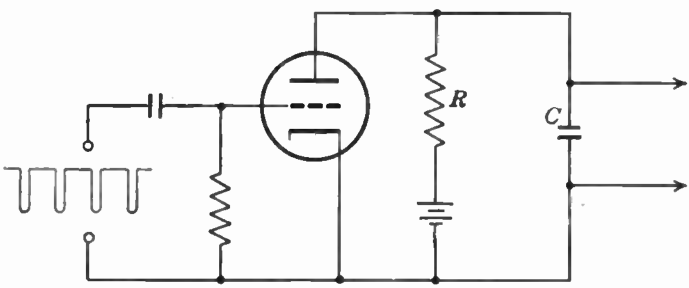


Fig. 4-27. Circuit for generating triangular pulses of voltage.

grid bias of the triode is normally zero. The small current passed by the resistor R flows through the tube, and the voltage of the condenser remains negligibly low because the plate voltage corresponding to this plate current is very low at zero grid voltage. The triode is periodically biased beyond cutoff by means of voltage pulses impressed upon the grid through the coupling circuit. During the duration of a pulse the condenser charges through the resistor R. If the maximum voltage to which the condenser charges is small in comparison with the supply voltage, or if the resistor is replaced by a pentode, the condenser voltage rises substantially linearly. In order to insure that the condenser starts and stops charging abruptly, the grid pulses should be rectangular in form. They may be obtained from a circuit of the form of Fig. 4-9. Figure 4-28a shows a typical wave of condenser voltage in the circuit of Fig. 4-27. The rate at which the condenser voltage rises is controlled by the sizes of the condenser and the resistance. The duration of the voltage pulses and the frequency of repetition is controlled by the

6. Haworth, L.J., Rev. Sci. Instr., 12, 478, Oct. 1941.

length and frequency of the rectangular pulses applied to the grid.

If the condenser charges so rapidly that it is fully charged before the end of a grid-excitation pulse, and the charging current is kept constant by the use of a pentode in place of the resistance R, the condenser voltage is of the form shown in Fig. 4-28b.

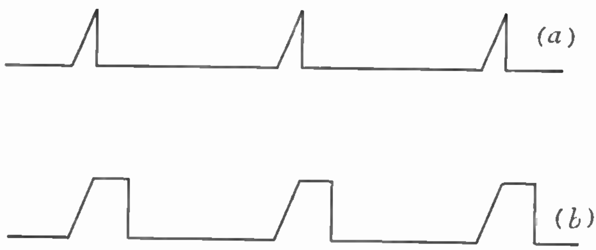


FIG. 4-28. Voltage waves generated by (*a*) the circuit of Fig. 4-27; (*b*) a similar circuit in which *R* is replaced by a pentode.

Figure 4-29 shows a complete circuit for producing voltages of the form shown in Fig. 4-28. This circuit is a combination of the circuits of Figs. 4-23, 4-10, 4-9, and 4-27. The functions of the various portions of the circuit are shown by the wave forms of the voltages at various points.

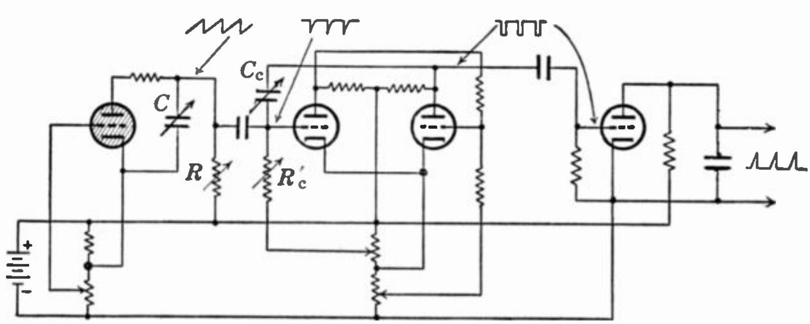


FIG. 4-29. Complete circuit for the generation of a voltage of the form of Fig. 4-28*a*.

4-16. **Sine-wave Oscillators.**- A general discussion of all types of oscillators for the production of sinusoidal voltages is beyond the scope of this book. The treatment that follows will therefore be limited to the types that are most frequently used in the production of audio- and radio-frequency oscillations in connection with ultra-high-frequency experiments. These are negative-resistance oscillators, including one type of RC (resistance-capacitance) or "resistance-tuned" oscillator, feed-back oscillators, beat-frequency, or heterodyne, oscillators, and bridge-type RC oscillators. Since feed-back oscillators can be treated as negative-resistance oscillators, the distinction between negative-resistance and feed-back oscillators is in a sense artificial. Under negative-resistance oscillators, however, will be considered only those types which make use of a circuit element that displays negative a-c resistance even when it is not used in connection with an oscillatory circuit.

Ultra-high-frequency generators, as distinguished from audio- and radio-frequency oscillators, are treated separately in Ch. 10.

4-17. Theory of Negative-resistance Oscillators.- Application of Kirchhoff's laws to the circuit of Fig. 4-30 gives the following equation for the current in any branch of the circuit:

$$\frac{d^2i}{dt^2} + \left(\frac{r}{L} + \frac{1}{\rho C}\right)\frac{di}{dt} + \frac{r+\rho}{L\rho C}\, i = 0 \qquad (4\text{-}7)$$

where r and ρ are the resistances, L the inductance and C the capacitance indicated in the figure, and all are assumed constant. Solution of this equation gives for the current:

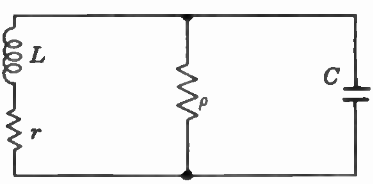


FIG. 4-30. Basic circuit of negative-resistance oscillator.

$$i = Ae^{-\frac{1}{2}(r/L + 1/\rho C)t}\sin\omega t \qquad (4\text{-}8)$$

in which A is a constant and the angular frequency of oscillation is

$$\omega = \sqrt{\frac{r+\rho}{\rho}\,\frac{1}{LC} - \frac{1}{4}\left(\frac{1}{\rho C} + \frac{r}{L}\right)^2} \qquad (4\text{-}9)$$

If ω is a real quantity, eq. (4-8) shows that the circuit currents vary sinusoidally at an amplitude that may decrease, remain constant, or increase. If the quantity $(r/L + 1/\rho C)$ is positive, the exponential factor in eq. (4-8) decreases with time, and so the amplitude decreases with time and oscillation eventually ceases. If the quantity $(r/L + 1/\rho C)$ is negative, on the other hand, the exponential factor increases with time and the amplitude of oscillation builds up. In the critical case in which $(r/L + 1/\rho C)$ is equal to zero, the exponential factor is unity, indicating that the amplitude of oscillation remains constant. Only if either r or ρ is negative, is it possible for $(r/L + 1/\rho C)$ to be negative or zero. In the critical case in which

$(r/L + 1/\rho C)$ is equal to zero, i.e., when $\rho = -L/rC$ and the amplitude of oscillation is constant, eq. (4-9) reduces to

$$\omega = \sqrt{\frac{r + \rho}{\rho} \frac{1}{LC}} \tag{4-10}$$

In practice, r is small in comparison with ρ, and so the frequency of oscillation is practically equal to $1/(2\pi\sqrt{LC})$.

A negative value of resistance r or ρ indicates that a positive increment of voltage across the element is accompanied by a negative increment of current, i.e., the slope of the current-voltage characteristic of the element is negative in the given current range. Several types of negative-resistance elements are available. In most negative-resistance oscillators the element ρ has negative resistance, and r is then the positive resistance of the inductance coil.

If ω is an imaginary, in contrast with the case considered in the preceding paragraphs, eq. (4-8) shows that the current may fall exponentially to zero, remain constant, or increase exponentially without limit, depending on the sign and magnitude of ω. Unlimited increase of current is, of course, a physical impossibility. Actually the current may in some cases rise exponentially to a critical value at which there is an abrupt change of some circuit parameter, and the current starts falling exponentially to a second critical value at which another abrupt change of a circuit parameter causes it to start rising again. As pointed out in Sec. 4-1, triggering may occur in a circuit containing a negative-resistance element. It is the triggering of the circuit that produces the abrupt changes in parameters necessary to cause the current to rise and fall periodically. The production of relaxation oscillations in circuits containing negative-resistance elements has already been discussed in Sec. 4-7, in which it was also pointed out that the currents in such circuits vary exponentially between the critical values at which triggering occurs. Experimental studies verify the predictions that relaxation oscillations should occur when the circuit constants are such as to make ω imaginary. An analysis of eq. (4-9) shows that relaxation oscillations should occur when the magnitude of $L/\rho rC$ is large. This prediction is verified experimentally. When the tuning capacitance is merely the small distributed capacitance of the inductance coil, for instance, the oscillations are not usually sinusoidal in form.

4-18. **Practical Negative-resistance Oscillators.**- Reference to Fig. 4-3 shows that the characteristic curve relating the screen current to the screen voltage of a pentode in which the screen is coupled to the suppressor has a portion with negative slope. Between the screen and the screen voltage supply, therefore, such a circuit may act like a negative resistance, and sustained sinusoidal oscillations may occur if a condenser in parallel with an inductance is inserted in the screen circuit. This type of oscillator is called the "transitron" or "negative-transconductance" oscillator. Figure 4-31 shows a typical circuit in which the suppressor is coupled to the screen through a condenser C_c the reactance of which is negligible in comparison with the resistance R_c at the frequency of the resonant "tank" circuit.

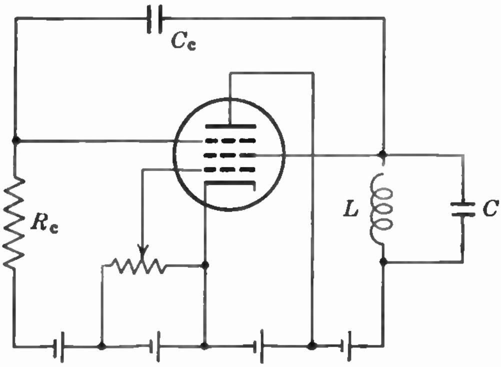


FIG. 4-31. Transitron oscillator.

Figure 4-6, which shows the characteristic relating external current between points A and B of the circuit of Fig. 4-5 with the impressed voltage that causes it to flow, indicates that a negative resistance exists between points A and B. Sustained sinusoidal oscillations should therefore be produced if a parallel combination of inductance and capacitance is connected between A and B. Figure 4-32 shows a practical negative-resistance oscillator based upon this circuit. The reactance of the coupling condensers C_c should be negligible in comparison with R_c at the frequency of the tank circuit.

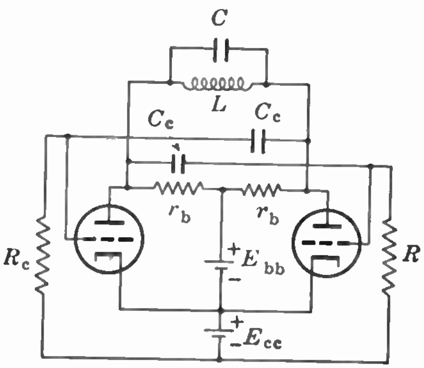


FIG. 4-32. Push-pull negative-resistance oscillator.

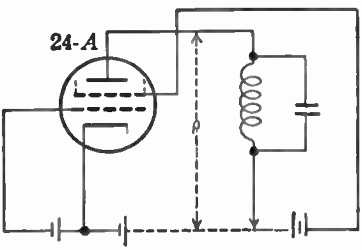


FIG. 4-33. Dynatron-oscillator circuit.

Figure 4-33 shows the circuit of a dynatron oscillator, which makes use of the negative screen resistance of a screen-grid tetrode caused by secondary emission from the plate (see Sec. 2-6). This type of oscillator is less reliable than the transitron and push-pull circuits, because of the fact that the secondary emission characteristics of the tube are likely to vary during the life of the tube.

The frequency of oscillation of the circuits of Figs. 4-31, 4-32, and 4-33 is very nearly equal to $1/2\pi\sqrt{LC}$. The amplitude of oscillation may be readily controlled by means of the control-grid bias. It is of interest to note the similarity between the transitron oscillator, the pentode trigger circuits, and the van der Pol relaxation oscillator. These three circuits are, in fact, all special forms of one circuit, and the mathematical analysis of Sec. 4-17 applies to all three. Similarly, the push-pull negative resistance oscillator, the Eccles-Jordan trigger circuit, and the multivibrator are special forms of a single circuit. The same is true of the dynatron, the tetrode trigger circuit, and a relaxation oscillator which may be based upon the tetrode circuit.

4-19. Harmonic Content of Negative-resistance Oscillators.- In order to prevent excessive harmonic generation, the amplitude of oscillation of a circuit containing a non-linear element must be kept small. The ease with which the amplitude of oscillation of a negative-resistance oscillator can be controlled depends upon the manner in which the average or dynamic negative resistance varies with amplitude. If the dynamic negative resistance decreases in magnitude with increase of amplitude the amplitude increases until it becomes so high that the negative resistance again decreases with further increase of amplitude. If, on the other hand, the dynamic negative resistance increases in magnitude with amplitude, and the static value of ρ is made slightly smaller in magnitude than L/rC in order to ensure starting of oscillations, the amplitude need build up by only a small amount to make the dynamic value of ρ equal to L/rC. The dynamic negative resistance increases in magnitude with amplitude of oscillation if the current-voltage characteristic of the negative-resistance element is of such shape that the operating point is at a point of inflection above and below which the slope of the curve is less than at the point. Because this may be true of the negative-resistance elements used in the circuits discussed in the preceding section, it is possible to maintain low amplitude of oscillation in these circuits and thus to make the harmonic content small.

The harmonic content is least when the output voltage is taken from across the tuned circuit or from a second inductance coupled to the tank inductance. When appreciable current or power is required by the load, the load is likely to affect the frequency or even to stop oscillation if the load is connected directly to the oscillator. For this reason, the oscillator is usually used to excite an amplifier, the output of which is delivered to the load.

Another advantage of negative-resistance oscillators over feed-back oscillators is their high frequency stability, which is made possible by the low amplitude of oscillation attainable. A third advantage is the ease with which the frequency range may be altered by changing the size of a single inductance.

4-20. Negative-resistance Oscillators Without Inductance.- By the use of inverse feedback, the amplification of the amplifier of Fig. 4-34 may be made independent of frequency throughout the frequency range in which it is used.[7] The output voltage is then in phase with or opposite in phase to the input voltage. The voltage amplification[8] is assumed to be positive when the input and output voltages are in phase, as indicated by the signs in Fig. 4-34. It is assumed that the input impedance of the amplifier is infinite, so that no current flows into the input of the amplifier.

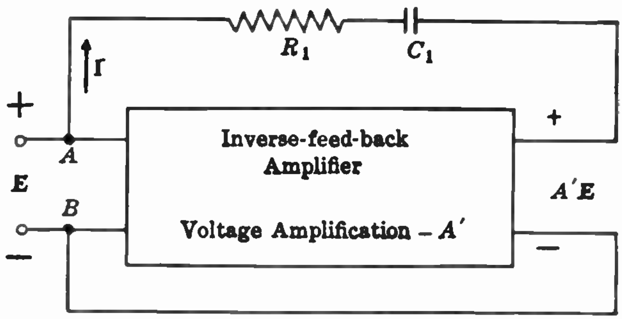


FIG. 4-34. Circuit for producing negative resistance.

The current $\underline{I}$ that flows[7] as the result of application of the voltage $\underline{E}$ between points A and B is

$$\underline{I} = \frac{\underline{E}(1 - \underline{A})}{R_1 + \frac{1}{j\omega C_1}} \qquad (4\text{-}11)$$

and the effective impedance between points A and B is

$$\underline{Z_e} = R_1/(1 - \underline{A}) + 1/(1 - \underline{A})j\omega C_1 \qquad (4\text{-}12)$$

If the output voltage is in phase with the input voltage, $\underline{A}$ is positive, and between points A and B the circuit acts like a negative resistance in series with an effective inductance of magnitudes

7. See Ch. 3.

8. Voltage amplification is discussed in detail in Ch. 3.

$$R_e = R_1/(1 - A) \qquad (4\text{-}13)$$

$$L_e = 1/(A - 1)\omega^2 C_1 \qquad (4\text{-}14)$$

If, therefore, a parallel combination of resistance R_2 and capacitance C_2 is connected between points A and B, the resulting circuit is of the form of that of Fig. 4-30, in which

$$r = R_e = R_1/(1 - A) \qquad (4\text{-}15)$$

$$L = L_e = 1/(A - 1)\omega^2 C_1 \qquad (4\text{-}16)$$

$$C = C_2 \qquad (4\text{-}17)$$

$$\rho = R_2 \qquad (4\text{-}18)$$

Since R_e is negative, sustained oscillations of constant amplitude are obtained when

$$r = L/\rho C \qquad (4\text{-}19)$$

Substitution of eqs. (4-15) to (4-18) in eqs. (4-10) and (4-19) shows that the criterion for oscillation and the frequency of oscillation are:

$$\underline{A} = \frac{R_1}{R_2} + \frac{C_2}{C_1} + 1 \qquad (4\text{-}20)$$

$$f = 1/2\pi\sqrt{R_1 R_2 C_1 C_2} \qquad (4\text{-}21)$$

Equation (4-21) shows that the frequency of oscillation can be controlled by means of R_1, R_2, C_1, or C_2 or by two or more of these parameters. If the resistances are equal and the capacitances are equal, eq. (4-21) reduces to

$$f = 1/2\pi R_1 C_1 \qquad (4\text{-}22)$$

The fact that the frequency may thus be made inversely proportional to the capacitance, instead of to the square root of the capacitance, makes it possible to cover a 10-to-1 frequency range with ganged condensers of the type used in broadcast receivers. The frequency band may be readily changed by changing R_1 and R_2. The circuit constants must, however, be chosen so that the oscillation is of the sinusoidal type.

Figure 4-35 shows a practical form of this circuit used in one of the most successful commercial resistance-capacitance oscillators.[9] The ballast lamp R_3, in combination with the resistance R_4, provides inverse feedback that makes the amplification and phase shift of the two-stage resistance-capacitance-coupled amplifier independent of input voltage, variable circuit parameters, supply voltage, and tube characteristics, and at the

9. Terman, F.E., Buss, R.R., Hewlett, W.R., and Cahill, F.C., Proc. I.R.E., 27, 649 (1939).

same time affords a method of stabilizing the amplitude of oscillation. Increase of amplitude raises the current through the lamp and thus increases its resistance and hence the inverse feedback. Increase of feedback decreases the voltage amplification of the amplifier and thus tends to reduce the amplitude of oscillation.

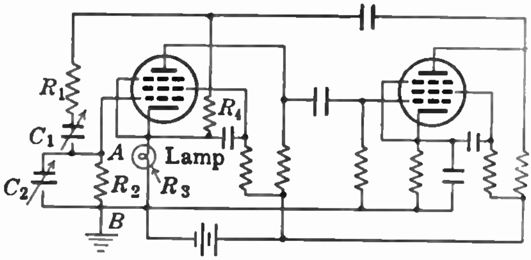


FIG. 4-35. Circuit diagram of resistance-tuned oscillator.

When a conventional resistance-coupled amplifier is used in the circuit of Fig. 4-34, an even number of stages is required in order to make the output voltage in phase with the input voltage. Because the mu-factor relating the suppressor and screen voltages of a pentode is negative, however, a single pentode may be used in this circuit, as shown in Fig. 4-36. The purpose of the resistance R_f and the condenser C_f is to provide inverse feedback in order to improve wave form and stability. The frequency is varied by means of C_1, C_2, and R_2. The condenser C_2 may shunt the screen resistor, instead of the suppressor resistor.

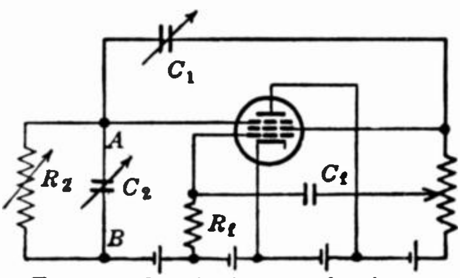


FIG. 4-36. Circuit diagram of resistance-tuned oscillator.

4-21. **Feed-back Oscillators.**- A feed-back oscillator can be considered as a tuned feed-back amplifier (see Ch. 3) in which the amplitude and phase angle of the feed-back voltage are such as to cause oscillation. Suppose that a voltage $\underline{E}_1$ is applied to the input of an amplifier and that the resulting output voltage is $\underline{E}_o$. If a portion of the output voltage is applied to the input, in addition to $\underline{E}_1$ and in phase with it, this feed-back voltage will act in the same manner as $\underline{E}_1$. If the magnitude of the feed-back voltage is equal to $\underline{E}_1$, it can replace $\underline{E}_1$ and the amplifier will continue to deliver the original output $\underline{E}_0$ when $\underline{E}_1$ is removed. This is another way of saying that the amplifier will oscillate at constant amplitude. If the feedback is increased, the amplitude will build up; if it is decreased, the amplitude will die down.

Figure 4-37 shows the generalized circuit for single-tube feed-back oscillators, from which the more common circuits can be derived by eliminating one or two of the condensers or by making M zero. The three most commonly used circuits, the tuned-grid, the tuned-plate, and the Hartley circuits, are shown in Figs. 4-38, 4-39, and 4-40. (In the Colpitts oscillator, not shown, C_3 is replaced by an inductor, and L_1 and L_2 are omitted.)

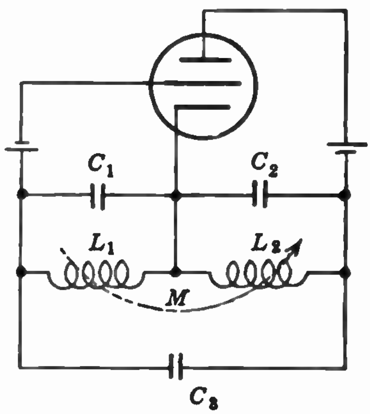


Fig. 4-37. Generalized feed-back oscillator circuit.

As in negative-resistance oscillators, the output voltage of feed-back oscillators may be taken from across the tank circuit or, preferably, from a separate coil coupled to the tank inductance.

Feed-back oscillators are usually analyzed by the use of equivalent plate circuits,[10] even though the amplitude of oscillation

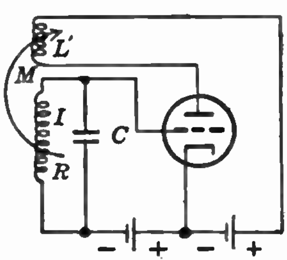


Fig. 4-38. Tuned-grid oscillator.

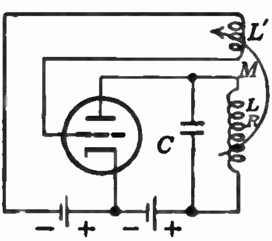


Fig. 4-39. Tuned-plate oscillator.

10. For this reason amplifier analysis, such as much of that of Ch. 3, and oscillator analysis are similar in many respects. Fundamentally the oscillator analysis using the equivalent plate circuit reduces to assuming an applied voltage in the grid circuit, and determining from the amplifier equations the conditions under which plate current would flow if the applied grid voltage were zero.

may be so high that the path of operation is far from linear. Although this method of analysis gives no indication of the production of harmonics, it yields a large amount of valuable information concerning the fundamental frequency of oscillation and the conditions which must be satisfied in order that sustained sinusoidal oscillations may be produced. In order to simplify the analysis, it is customary to neglect the effect of grid current. Although the flow of grid current that may occur during the positive peaks of grid voltage cannot be neglected in its actual effect on oscillator operation, it is usually satisfactory in an approximate analysis to make use of the simpler assumption of no grid current and to bear in mind that the results may have to be modified to take the effects of grid current into account. One effect of the flow of grid current is to introduce damping or power loss, which may be considered as equivalent to an increase of the resistance of the tuned circuit. Grid current may increase harmonic content, since reduction of plate current resulting from diversion of electrons to the grid increases the curvature of the dynamic transfer characteristic. It will be seen, however, that the flow of grid current may be used advantageously in limiting the amplitude of oscillation by automatically increasing the grid bias with amplitude.

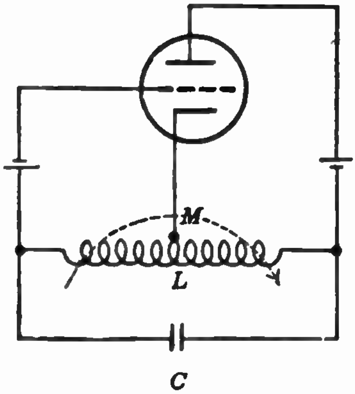


FIG. 4-40. Hartley oscillator.

Figure 4-41 shows the equivalent plate circuit for the tuned-plate oscillator under the assumption that condenser losses are negligible. The alternating grid voltage $\underline{E}_g$ is induced in the grid coil by virtue of magnetic coupling to the plate coil. If M is assumed to be positive when an increase of $\underline{I}$ results in a positive induced voltage, $\underline{E}_g = j\omega M\underline{I}$. Simultaneous solution of this equation with two equations obtained by summing voltages in the equivalent circuit gives the following equations for the criterion for oscillation and the frequency of oscillation:

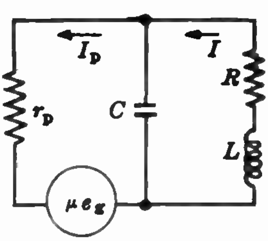


FIG. 4-41. Equivalent plate circuit for tuned-plate oscillator.

$$g_m \geq |rC/M + L/Mr_p| \qquad (4\text{-}23)$$

$$f = \frac{1}{2\pi\sqrt{LC}}\sqrt{\frac{r + r_p}{r_p}} \qquad (4\text{-}24)$$

When no power is drawn from the oscillating circuit, the effective resistance r is small in comparison with r_p, and the frequency of oscillation is practically the natural frequency of the resonant circuit, $1/2\pi\sqrt{LC}$. If g_m exceeds the value given by eq. (4-23), the amplitude of oscillation increases with time. From eq. (4-23) and the expression for the induced alternating grid voltage, it can be seen that sustained oscillation can take place only when the coupling is of such sign as to result in a positive increment of grid voltage when the current through the plate inductance increases.

Similar analyses for the tuned-grid oscillator and the Hartley oscillator show that the frequency of oscillation of these circuits also approximates $1/2\pi\sqrt{LC}$, and that the transconductance must exceed a critical value, which varies with coupling, in order for sustained oscillations to be possible.

The form of eq. (4-23) and of similar equations derived for other feed-back oscillator circuits shows clearly the importance of high transconductance in tubes used in feed-back oscillators. These equations also show that, for a given transconductance, oscillation will take place more readily the higher the plate resistance. Since at a given transconductance the amplification factor is proportional to the plate resistance, it follows that the "figure of merit" of a tube for use in a feed-back oscillator is the product μg_m. As this is also the figure of merit of a power tube, power tubes are in general good feed-back oscillator tubes.

4-22. Bridge-type Feed-back Oscillator.- Figure 4-42 shows the block diagram of a feed-back oscillator which is tuned

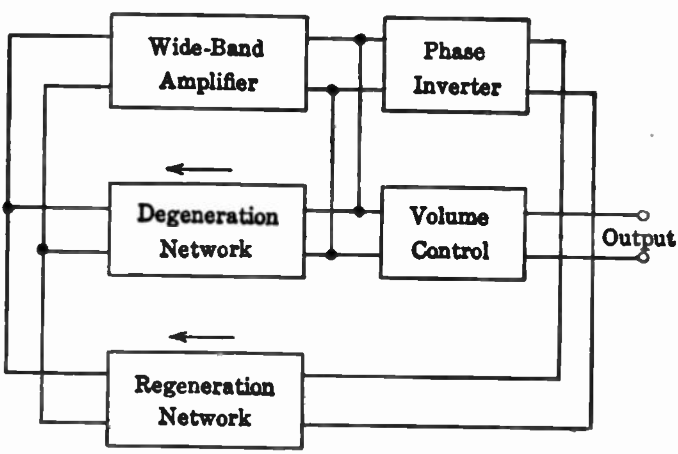


FIG. 4-42. Block diagram of feed-back R-C oscillator.

by means of a resistance-capacitance bridge. The amplifier is designed to have constant amplification and zero phase shift over a wide range of frequency. Degenerative feedback is applied to the amplifier through an impedance bridge. Since the output of the bridge is zero at resonance, the inverse feedback is a minimum at this frequency and the amplification of the amplifier with inverse feedback is a maximum. The response curve of the amplifier has a sharp peak at the resonance frequency of the bridge. The regenerative-feed-back network, which has a flat response throughout the frequency range of the amplifier, applies positive feedback to the amplifier. If the positive feedback is high enough, sustained oscillation takes place at the frequency at which the amplification is greatest, i.e., at the resonance frequency of the bridge. The phase inverter is necessary in order to make the feed-back voltage of correct phase to give regenerative feedback. A satisfactory form of degenerative-feed-back bridge is the Wien bridge, shown in Fig. 4-43, which contains only resistances and capacitances. If $R_3 = 2R_4$, $R_1 = R_2$, and $C_1 = C_2$, the bridge is balanced when $f = 1/2\pi R_1C_1$, and the circuit oscillates at this frequency. This type of oscillator has the same advantages as the resistance-capacitance negative-resistance oscillator discussed in Sec. 4-20. It may be tuned either by ganged variable condensers or ganged variable resistors.

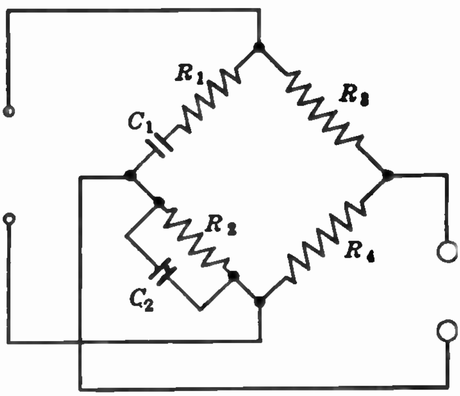


FIG. 4-43. Wien bridge.

4-23. Push-pull Oscillators.- The basic feed-back oscillator circuits may be modified to use two tubes in push-pull. As in amplifiers, the use of push-pull circuits increases the power output and decreases the harmonic content. The frequency stability of push-pull circuits is also higher than that of single-sided circuits. They are used principally at high and ultra-high frequencies.

Circuits in which the plate-supply voltage is connected in series with the plate inductance, as in Figs. 4-38, 4-39, and 4-40, are called series-feed circuits. In practice it may be desirable or necessary to connect the plate to the oscillating circuit through a condenser of low reactance and to apply the direct voltage to the plate through a choke, the reactance of which is so high that it does not appreciably affect the oscillating circuit. Air-core chokes are used in radio-frequency circuits, iron-core chokes in audio-frequency circuits. An example of this method of applying the direct plate voltage, called parallel feed, is given by the circuit of Fig. 4-44b. The objection to the use of parallel feed is that parasitic oscillations may take place in the feed condenser and choke.

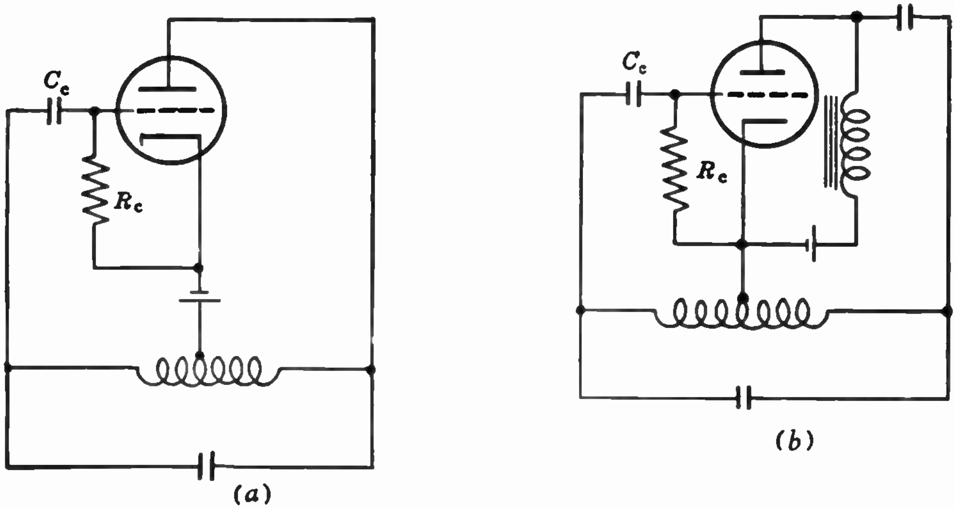


FIG. 4-44. Use of grid condenser and leak to limit amplitude of oscillation. Hartley oscillator with (a) series feed, (b) parallel feed.

4-24. **Use of Self-bias to Limit Amplitude of Oscillation.**- For the purpose of simplicity of representation, fixed biasing voltages have been indicated in the basic circuits of Figs. 4-37 to 4-40. Fixed bias is rarely used, however, in practical oscillators. In order to prevent excessive distortion and to aid in obtaining frequency stabilization, it is necessary to limit the amplitude of oscillation. In feed-back oscillators the criterion for oscillation involves the transconductance of the tube, and the amplitude builds up until the average (dynamic) transconductance drops to the critical value below which oscillation cannot take place. Unfortunately the average transconductance first increases with amplitude, and so the amplitude may increase to a high value before the average transconductance again falls sufficiently to result in equilibrium. If the circuit is adjusted to give small equilibrium amplitude, then it will not start of its own accord. A similar difficulty may be experienced with some negative-resistance oscillators. This difficulty may be prevented by causing the grid bias to increase automatically with amplitude.

The most common method of limiting the amplitude of oscillation is the use of a grid-blocking condenser and grid leak, as shown in Fig. 4-44. The initial bias is zero, but, as soon as oscillation commences, the grid is driven positive during a portion of the cycle and so electrons flow from the cathode to the grid. During the remainder of the cycle these electrons cannot return to the cathode but can only leak off the condenser and grid through the grid leak R_c. The trapped electrons make the potential of the grid negative with respect to the cathode, thus providing a bias. The bias may also be considered to result from the flow of current through the grid leak. The greater the amplitude of oscillation, the more positive the grid swings, and the greater is the average grid current. Thus the bias builds up with oscillation amplitude, causing the transconductance to fall until equilibrium is established. In this manner the amplitude may be prevented from becoming too high without making the quiescent transconductance so low as to prevent oscillation from starting spontaneously. Under equilibrium conditions grid current flows during only a very small fraction of the cycle, and the grid bias is very nearly equal to the amplitude of the alternating grid voltage.

Low power loss in the resistor, high frequency stability, and good wave form call for the use of high grid-leak resistance; but it is found that, if the resistance is too high, oscillation is not continuous. After a number of cycles of oscillation the bias becomes so high that the circuit stops oscillating. Because oscillation starts at a lower bias than that at which it stops, some time elapses while the condenser discharges sufficiently to allow oscillation to recommence, and so periods of oscillation alternate with periods of rest. This phenomenon is termed motorboating. The period of motorboating depends upon the time required for the condenser to discharge, which increases with the product of the grid condenser capacitance and the grid-leak resistance.

Another factor that limits the size of the grid resistor is danger of cumulative increase of positive grid voltage and plate current as the result of primary and secondary grid emission.[11]

4-25. Frequency Stability.- Undesired changes of frequency result from three major causes: changes in the mechanical arrangement of the elements of the oscillating circuit; in the values of the circuit parameters; and in the amplification factor, grid and plate resistances, and interelectrode capacitances of the tube. Changes in the mechanical arrangement of the circuit elements may be produced by vibration; by mechanical, electric, or magnetic forces; or by temperature changes. They can

11. Reich, H.J., "Theory and Applications of Electron Tubes," pp. 184-185.

be minimized by careful mechanical and electrical design and by temperature control.

Variations in the values of the circuit parameters result from changes in temperature of inductances and condensers and from variation of load, which alters the effective a-c resistance of the tuned circuit. Changes of inductance and capacitance can be minimized by: temperature control; the use of thermally compensated inductances and temperature-controlled compensating condensers; and the careful choice of apparatus and the judicious location of component parts. The most common method of preventing the load from affecting the frequency is to take the required power from a "buffer" amplifier that is excited by the oscillator, rather than from the oscillator directly.

Tube factors and electrode capacitances are dependent upon operating voltages, upon cathode emission, and upon electrode spacing. Operating voltages can be stabilized by the use of voltage-regulating devices. Variation of cathode emission is probably the least important factor and can be reduced by the maintenance of rated cathode temperature. Electrode spacing, which depends to some extent upon tube temperature, affects the interelectrode capacitances. The dependence of frequency upon interelectrode capacitances and upon stray circuit capacitance can be minimized by the use of a high ratio of tuning cpacitance to inductance and by the use of circuits in which the tuning capacitance shunts the grid-plate capacitance.

Dependence of frequency upon plate resistance may be minimized by the use of resistance-stabilized circuits, in which a high resistance in series with the plate reduces the percentage change of total plate-circuit resistance.[12] The most effective method of frequency stabilization, however, is the use of vibrating mechanical elements electrostatically or magnetically coupled to the oscillating circuit. At radio frequencies this is accomplished by the use of quartz crystals, in which compression or expansion results in the production of a potential difference between opposite faces and, conversely, application of a potential difference results in an elongation or contraction.[13] At audio frequencies it is accomplished by the use of magnetostrictive rods, which expand or contract when magnetized.[14] In quartz-crystal oscillators the frequency variation may be made

12. Horton, J.W., Bell System Tech. J., 3, 508 (1924); Terman, F.E., Electronics, July, 1933, p. 190.
13. Cady, W.G., Proc. I.R.E., 10, 83 (1922). For additional references see H.J. Reich, "Theory and Applications of Electron Tubes," p. 335.
14. Pierce, G.W., Proc. I.R.E., 17, 42 (1929); Salisbury, W.W., and Porter, C.W., Rev. Sci. Inst. 10, 142 (1939).

less than two parts in a million; in magnetostriction oscillators it may be limited to one part in 30,000.

4-26. Beat-frequency (Heterodyne) Oscillators.[15]- An entirely different type of audio-frequency oscillator is the beat-frequency oscillator. In an oscillator of this type, shown schematically in Fig. 4-45, the outputs of two radio-frequency oscillators of slightly different frequencies are applied simultaneously to a detector. The output of the detector contains,

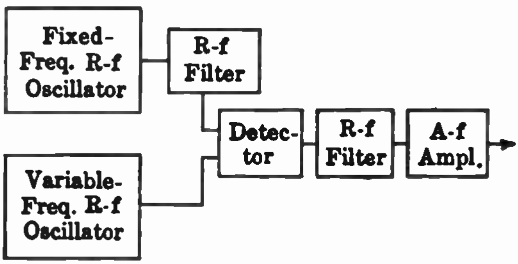


FIG. 4-45. Schematic diagram of heterodyne oscillator.

in addition to the impressed radio frequencies, their sum and difference. By means of a filter the fundamental radio frequencies and their sum are removed, leaving only the difference frequency in the output, which may be amplified by audio-frequency amplifiers. The popularity of the heterodyne oscillator is due principally to the fact that the whole range of audio frequencies, from 15,000 cycles, or higher, down to as low as one cycle, may be covered with a single dial. Other advantages that may be obtained with careful design include good wave form, constant output level, lightness, and compactness. By proper variable condenser design a logarithmic frequency scale may be obtained, a considerable advantage when the oscillator is to be used in obtaining amplifier-response curves. Unless extreme care is taken in the design and construction, however, this type of oscillator is likely to have relatively poor frequency stability, which necessitates frequent setting against a standard frequency during the period of use, particularly during the time required to establish temperature equilibrium.

4-27. Crystal Oscillators.- The most satisfactory method of stabilizing the frequency of radio-frequency oscillators is by the use of quartz crystals. The control of frequency by means of crystals is based upon the piezoelectric effect. When certain crystals, notably quartz, are compressed or stretched in

15. See bibliography, H.J. Reich, "Theory and Applications of Electron Tubes." p. 364.

certain directions, electric charges appear on the surfaces of the crystal that are perpendicular to the axis of strain. Conversely, when such crystals are placed between two metallic surfaces between which a difference of potential exists, the crystals expand or contract. If the potential applied to the plates is alternating, the crystal is set into vibration, the amplitude being greatest at the mechanical resonance frequency of compressional oscillation of the crystal. Although special types of crystals may be used in a-f oscillators, crystal control is at present restricted mainly to r-f oscillators.

Figure 4-46 shows one commonly used crystal oscillator, developed by J. M. Miller at the Bureau of Standards. The action

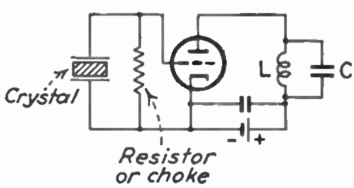


FIG. 4-46. One type of crystal-controlled r-f oscillator.

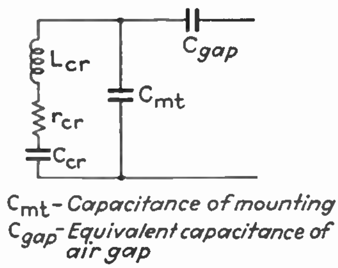


FIG. 4-47. Equivalent circuit of crystal and crystal mounting.

of this circuit is readily understood when it is noted that the crystal and crystal mounting may be represented by the equivalent electrical circuit of Fig. 4-47. When the plate load is inductive, the effective input conductance of the tube is negative. Therefore oscillation may be set up in a resonant circuit connected between the grid and cathode. In order to maintain an inductive plate load the plate circuit must be tuned so that its resonant frequency is slightly higher than that of the crystal. The value of the crystal in controlling frequency lies in the extreme sharpness of its resonance curve. The Q of the equivalent circuit of a crystal is of the order of one hundred times that which can be readily attained in electrical circuits. Because of this sharpness of resonance, the crystal can oscillate only over a very narrow frequency range, and hence the frequency stability of a crystal oscillator is high. When the temperature of the crystal is maintained constant by means of a temperature-control chamber, the frequency drift may be made less than two parts in 10 million.

In one of the early circuits for a crystal oscillator, the crystal was connected between the grid and plate of the tube, instead of between the grid and cathode. This circuit will oscillate only when the plate load is capacitative, and therefore the natural frequency of the plate circuit must be slightly lower than that of the crystal. A crystal will also oscillate when connected to other types of negative-resistance elements, such as were described earlier in this chapter. Crystals may also be used in push-pull circuits.

Because of the extremely sharp resonance of a crystal, it may also be used as a band-pass filter which passes a very narrow band of frequencies. One application of such a filter is in wave analyzers.

Chapter 5

CATHODE-RAY TUBES AND CIRCUITS

The cathode-ray oscillograph, which was originally developed for the study of voltage and current wave form, has found a great many other applications. Included in these are the measurement of voltage, time, and frequency, and the determination of characteristic curves relating to variable quantities such as electrode current and voltage of an electron tube. The heart of the cathode-ray oscillograph is the cathode-ray tube itself, which is also an essential component in television.

5-1. **The Cathode-ray Tube.**- Cathode-ray tubes consist of three essential parts: an electron gun for producing a narrow beam of rapidly moving electrons, called cathode rays; a fluorescent screen which produces a luminous spot as the result of impact of the cathode rays; and means for deflecting the beam and hence displacing the spot as the result of currents or voltages applied to the deflecting mechanism.

The electron gun comprises a thermionic source of electrons, a grid for controlling the electron density of the beam, and hence the brightness of the luminous spot, and means for focusing the electron beam sc that a small symmetrical spot is produced upon the screen. The focusing may be accomplished by means of either electric or magnetic fields. A discussion of the theory of electron optics is beyond the scope of this book, and the interested student is therefore referred to texts devoted to this subject.[1] It is sufficient here to state that it has been proved by rigorous electrodynamic analyses that electric and magnetic fields produce effects upon beams of electrons analogous to that of glass lenses upon light. Figure 5-1 shows a typical cathode-ray tube of the type ordinarily used in cathode-ray oscillographs. Electrostatic focusing is ordinarily used in such tubes. The grid, in addition to serving as a means of controlling the brightness of the fluorescent spot, also serves, in conjunction with the first anode, as a lens for focusing the electron beam. Electrons having velocity components normal to the axis of the tube are intercepted by the grid and by the diaphragms in the first anode. The electric field between the first anode and the second anode, which is of larger diameter and at a higher potential than the first anode, acts in a manner analagous to a converging lens and thus

---

1. See bibliography at end of chapter.

brings the electrons to a sharp focus on the screen. The final velocity of the electrons is determined by the potential of the

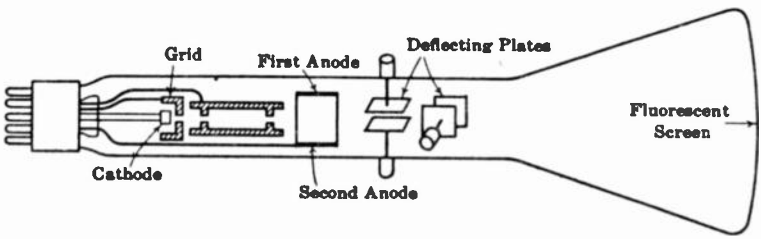


FIG. 5-1. Electrode structure of a typical cathode-ray oscillograph tube.

second anode, which is normally maintained constant in the operation of the tube. Focusing is accomplished by varying the potential of the first anode. The second anode sometimes consists of a conducting coating on the inside wall of the tube. It is of interest to note that the image formed on the screen may be that of either an aperture or of a concentration of electrons formed by the crossing of electron paths at a point a short distance beyond the grid. The designer of cathode-ray tubes is faced with a number of problems involving the electron gun, among which are departure of the beam from circular section and unequal focal distances for the various portions of the beam, with resultant lack of sharpness of the spot.

The electron beam may be focused magnetically by the use of either a uniform axial field extending the entire length of the tube or of an axially symmetrical non-uniform field produced by a short concentrated coil, the axis of which coincides with that of the tube. Because of the difficulty of producing a uniform axial field, magnetic focusing is ordinarily accomplished by means of one or more short coils.

5-2. Deflection of Electron Beam.- Deflection of the cathode-ray beam may be accomplished either by means of an electric field produced by two electrodes between which the deflection voltage is impressed, or by means of a magnetic field produced by electromagnets energized by deflection current. Figure 5-1 shows the arrangement of deflecting electrodes in a tube in which both vertical and horizontal deflection is produced by electrostatic fields. Since the field is normal to the surfaces of the deflecting electrodes, and the movement of electrons as the result of the field is parallel to the field, the deflection of the beam is in a plane normal to the electrode surfaces. A relatively simple electrodynamic analysis snows that the deflection of the luminescent spot from its normal position by parallel deflecting plates is equal to[2]

$$y = E_d ehL/v^2 md = E_d hL/2E_a d \qquad (5\text{-}1)$$

in which E_d is the voltage impressed between the deflecting plates, e is the charge of an electron in e.s.u., h is the length of the deflecting plates, L is the distance from the center of the deflecting plates to the screen, v is the electron velocity, m is the mass of an electron, d is the separation of the deflecting plates, and E_a is the potential of the second anode relative to the cathode. In the derivation of eq. (5-1) it is assumed that the field is uniform throughout the distance d and zero elsewhere. [In (5-1) E_d and E_a should be in same units; h, L, d, y likewise.]

It is evident from eq. 5-1 that the deflection of the luminescent spot is directly proportional to the deflecting voltage, and is inversely proportional to the voltage applied to the second anode. Since the brightness of the luminescent spot increases with the energy with which the electrons strike the screen, which is proportional to the second anode voltage, it follows that high values of brightness are attained at the expense of deflection sensitivity. This difficulty can be avoided by the use of a third anode, usually in the form of a grid, adjacent to the screen. The electrons may then be deflected while their velocity is low, and the high velocity essential to high spot brilliance may be given the electrons subsequent to deflection. Some defocusing is likely to result from this method, however. The deflection sensitivity may be increased by lengthening the plates, but the maximum deflection that can be obtained without having the electrons strike the deflecting plates is decreased by increase of plate length. For this reason deflecting plates are sometimes made parallel during a portion of their length, and divergent during the remainder, as shown in Fig. 5-2.

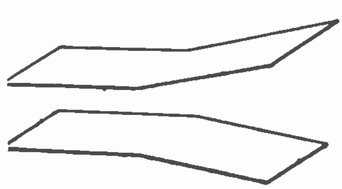


FIG. 5-2. Deflecting plates.

In many applications of cathode-ray tubes the spot is deflected by means of alternating voltages applied to the deflecting plates If the period of the deflecting voltage is of the order of magnitude of the time taken for the electrons to pass between the plates, the deflection of the spot is not proportional to the instantaneous voltage but to the integral of the voltage throughout

---

[2] See, for instance, H.J. Reich, "Principles of Electron Tubes," McGraw-Hill, 1941, Sec. 1-20.

the transit time between the deflecting plates. The high-frequency limit of the deflecting voltage that can be used in the study of voltage waveform is therefore dependent upon the transit time, which depends in turn upon the length of the deflecting electrodes and the second anode voltage. Tubes to be operated at high frequency must have relatively short deflecting electrodes and must operate at high voltage. Consequently such tubes have low deflection sensitivity. The limiting frequency at which cathode-ray tubes may be operated is also determined in part by the capacitance between the deflecting electrodes and their leads and by lead inductance. Capacitance and inductance are minimized by bringing the leads directly through the wall of the tube, as in Fig. 5-1, but for the sake of simplicity of manufacture and convenience of making connections, the leads are brought out through the base in most tubes designed for oscillographic use. "Crosstalk" between the vertical and horizontal deflecting plates is another problem that confronts the designer and user of cathode-ray tubes. It can usually be eliminated by proper circuit design.

Figures 5-3a and 5-3b show two coil arrangements suitable for producing electromagnetic deflection. In the arrangement of Fig. 5-3a the coils are connected so that the flux pro-

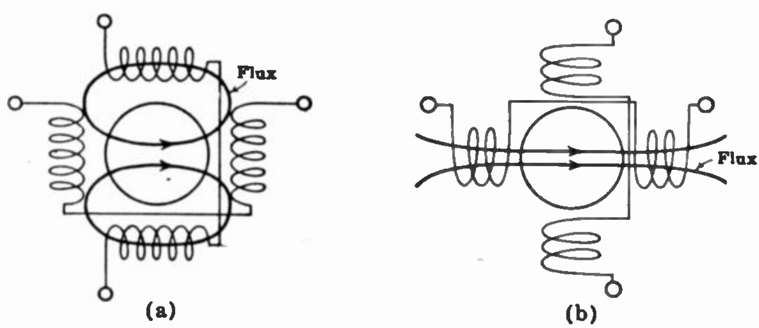
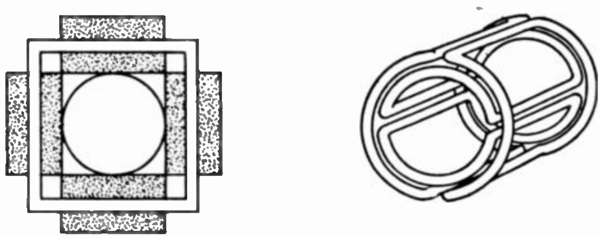


FIG. 5-3. Flux produced by two types of deflecting coils.

duced by the two coils of a pair bends in such a way as to combine in the region near the center of the tube. This arrangement produces a nearly linear field midway between the two coils. The method of winding and placing the coils is shown in greater detail in Fig. 5-4a. Because of the large air gap and the low magnetizing force that is ordinarily used, an iron yoke does not greatly increase the sensitivity, but may be useful in shaping the field. The coils may be conveniently held in place by mounting them on a wooden form, when an iron yoke is not used. In the arrangement of Fig. 5-3b, the coils are connected so that the flux produced by one coil of each pair adds to that of the other. This produces a nearly linear field parallel to the axis of the coil pair at a point midway between the coils. The coils may be



(a) (b)

FIG. 5-4. Structure of two types of deflecting coils.

bunch wound and wrapped about the neck of the cathode-ray tube, as shown in Fig. 5-4b. They may also be wound in a slotted core, as in a two-phase induction motor stator. To prevent crosstalk between the two pairs of coils, the coils of each pair should be as nearly alike as possible and should be symmetrically mounted relative to the other pair. To prevent introduction of deflection voltage into the focusing coils, a magnetic shield should be used between the focusing and deflecting coils.

The deflection produced by a magnetic field is normal to the direction of the field. Electrodynamic analysis indicates that the deflection is given by the expression:[3]

$$y = 0.297\ LsB/\sqrt{E_a}\ \text{cm} \qquad (5\text{-}2)$$

in which L is the distance in cm between the screen and the point where the electrons enter the field, s is the distance in cm over which the deflecting field acts, B is the flux density in gauss (assumed to be constant throughout the distance s), and E_a is the second anode voltage. Equation (5-2) shows that the deflection is proportional to the field strength, which is in turn proportional to the coil current when iron is not used. It is of interest to note that, although the deflection varies inversely with anode voltage, increase of anode voltage to raise the brightness of the spot does not decrease the sensitivity as rapidly as when electrostatic deflection is used. It has been shown that there is no advantage in making s, which is approximately equal to the length of the coils along the axis of the tube, greater than about 30 per cent of the distance L.[4]

3. See, for instance, Zworykin and Morton, "Television," Wiley, 1940, p. 449.
4. Maloff and Epstein, "Electron Optics in Television," McGraw-Hill, 1938, p. 195.

Although the deflection sensitivity in centimeters per ampere can be increased by increasing the number of turns of the coils, the resulting increase of inductance may be objectionable when high-frequency currents are used in the deflecting-coil circuits.

Cathode-ray tubes that are designed for electromagnetic deflection usually make use of magnetic focusing, whereas tubes designed for electrostatic deflection make use of electrostatic focusing. Sometimes, however, deflection in one direction is accomplished by electrostatic deflecting plates and at right angles by a deflecting yoke. The advantages of electromagnetic tubes over electrostatic tubes are the greater structural simplicity, which results in lower cost; the greater ruggedness, which makes for greater reliability in mobile equipment; and the shorter tube length, which reduces the overall size of the equipment in which they are used. Electrostatic tubes, on the other hand, require little or no deflection current or power. The auxiliary circuits are therefore simpler, as will be seen in a later section, and difficulties arising from deflecting-coil inductance are avoided. At the present time most cathode-ray oscillographs employ tubes with electrostatic focusing and deflection.

5-3. **Cathode-ray Tube Screens.**- The screen of a cathode-ray tube consists of a layer, of the order of 0.002 mm in thickness, of luminescent material applied to the end of the tube by spraying, dusting, or settling from a liquid. The bombardment of the screen by rapidly moving electrons gives rise both to fluorescence or emission of light during bombardment, and to phosphorescence or emission of light after bombardment. Choice of materials used in a screen designed for a particular type of tube is governed by the required luminous intensity of the spot per watt of the incident beam, by the color of the emitted light, and by the rate at which phosphorescence decays. Phosphors commonly used include willemite (zinc orthosilicate), which emits predominantly green light; zinc oxide, which emits predominantly blue light; zinc beryllium silicate, which emits predominantly yellow light; and a mixture of zinc sulphide with cadmium zinc sulphide or zinc berillium silicate, which emits nearly white light. Impurities must be present in these materials to act as "activators." Choice of color is influenced by whether the tube is to be used visually or photographically and by the purpose for which the tube is to be used. Slow decay of phosphorescence is desirable in a tube used for the observation of non-repeating phenomena or periodic phenomena having a low repetition frequency. Slow decay results in blurring, however, when the form of the pattern under observation changes, as in television reception or the study of voltages and currents of changing wave form. It is usually necessary to compromise between absence of blurring and freedom from flicker.

Continuous impact of electrons upon one spot of the screen reduces the sensitivity of the phosphor at that spot and may destroy the phosphor if allowed to take place over a long period. When high acceleration voltages are used, the heating may also be sufficient to melt the tube envelope. For this reason it is important to turn off the beam by applying a high negative grid bias whenever the beam is stationary. It is also unwise to operate the tube for a long time with the same pattern, particularly at high intensity. Most oscilloscopes are now provided with a convenient control for turning off the beam without turning off the cathode heater current. The student should form the habit of using this control and of working with as low brilliance as possible.

In the use of cathode-ray tubes, the objectionable effects of reflection of room light from the end of the tube can be reduced by the use of a filter of the same color as the light emitted by the screen.

5-4. Lissajous Figures.- Since the pattern produced on the screen of a cathode-ray tube by the application of voltages to the deflecting plates is usually substantially identical in form with the pattern produced by the flow of currents of the same wave forms through deflecting coils, it is sufficient to discuss the results obtained by electrostatic deflection. When direct voltages are applied simultaneously to the two pairs of deflecting plates the deflection is the resultant of those produced when the two voltages are applied individually. The application of alternating voltage to one set of plates causes the spot to trace a straight line on the screen. When alternating voltages are applied to both sets of plates the spot traces a complicated path that does not in general close if the frequencies of the two voltages are not the same, and that is therefore seen as a moving pattern commonly called a Lissajous figure. If the ratio of the two frequencies is a rational number, the path closes and the pattern is stationary. In the simplest case, in which the two frequencies are equal and the voltages pure sinusoids, the pattern may be a circle, an ellipse, or a straight line, depending upon the relative magnitudes and the phase difference between the two voltages. In general, a rational frequency ratio can be determined by enclosing the pattern with a rectangle the sides of which are parallel to the X and Y axes and tangent to the pattern. The ratio of the X to the Y frequency is equal to the number of points of tangency of the pattern to a vertical side divided by the number of points of tangency to a horizontal side. Figure 5-5 shows the Lissajous figure for a 3:2 frequency ratio.

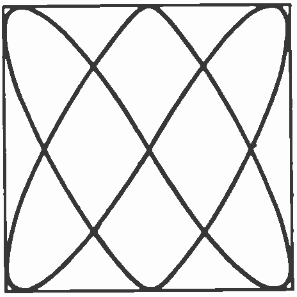


Fig. 5-5. Lissajous figure for 3:2 frequency ratio, enclosed by a rectangle.

5-5. Linear Sweep.- Although the wave form of one voltage may be readily determined from a Lissajous figure if the wave form of the other voltage is sinusoidal, it is generally far more convenient to see the unknown voltage as a function of time. This can be accomplished by applying to the horizontal deflecting plates a periodic voltage that increases linearly with time up to a maximum value and then falls abruptly to zero. This causes the spot to "sweep" across the screen at a uniform velocity, so that the deflection at any instant is proportional to the time elapsed since the beginning of the sweep. It is evident that the required voltage is saw-tooth in form and may be obtained from one of the oscillators discussed in Sec. 4-11 to 4-14. The arc-tube relaxation oscillator is used for this purpose in most oscilloscopes. The number of waves of Y voltage seen on the screen is equal to the ratio of the frequency of the Y voltage to the frequency of the sweep voltage. The pattern is stationary when the ratio is a rational number. Although the sweep frequency can be synchronized to the frequency of the voltage under observation by adjusting the sweep frequency, small changes in frequency of either voltage cause the pattern to drift. This difficulty is prevented by introducing a small amount of the Y voltage into the circuit of the sweep oscillator in a manner discussed in Sec. 4-10. "Locking-in" is accomplished with the least synchronizing voltage and hence with the least danger of distortion of the pattern when the free frequency of the sweep oscillator is slightly lower than a multiple of the frequency of the voltage under observation.

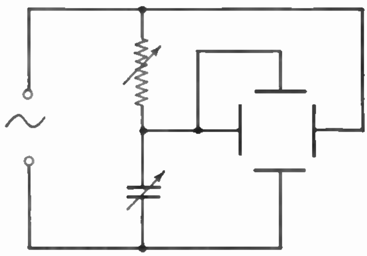
In the voltage produced by saw-tooth voltage oscillators the fall in voltage at the end of the linear rise in voltage takes up a portion of the cycle that may be appreciable at the higher frequencies of oscillation. The return of the spot at the end of the linear sweep is therefore not instantaneous, and a portion of the cycle of the wave under observation is traced during the return sweep. Although this portion of the pattern is of relatively low brightness, it may be objectionable, particularly at high frequencies. It may be eliminated by biasing the grid of the cathode-ray tube beyond cutoff during the return sweep. One method of obtaining the required pulse of negative voltage is from a small resistor in series with the tube that discharges the sweep-oscillator condenser, as, for instance, the current-limiting resistor used in series with the arc-discharge tube in the circuit of Fig. 4-23. This voltage may be amplified if necessary, and applied to the oscillograph grid through resistance-capacitance coupling. Another method of obtaining the pulse is by applying the sweep voltage to a condenser-resistance pulse-sharpening circuit of the type discussed in Sec. 4-6.

At sweep frequencies above 50 kilocycles the de-ionization time of the arc-discharge tube prevents the use of the simple arc-tube relaxation oscillator. One of the high-vacuum-tube relaxation oscillators discussed in Secs. 4-11 and 4-14 must then be used.

5-6. Linear Sweep with Long Rest Period.- It is sometimes necessary to use the cathode-ray oscilloscope in the observation of the wave form of a periodic phenomenon that is repeated at a relatively low frequency, but takes place in a very small fraction of the repetition period. If the ordinary type of sweep voltage is used, the portion of the pattern that is of interest occupies only a small portion of the screen, most of the pattern consisting of a horizontal line. This difficulty can be avoided by allowing the sweep to occur only during the time in which the phenomenon takes place. This is accomplished by using the circuit of Fig. 4-27, which gives the voltage shown in Fig. 4-28a. The spot is deflected only during the portion of the cycle in which the triangular pulse occurs. If the oscillator is synchronized to the phenomenon under observation in such a manner that the sweep is initiated at or slightly prior to the beginning of the phenomenon, the desired portion of the wave takes up the entire useful portion of the screen. Since the spot is stationary on the screen during the portion of the repetition cycle elapsing between two repetitions of the phenomenon, the grid of the cathode-ray tube must be biased beyond cutoff during this time.

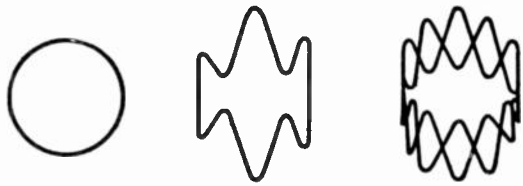
5-7. Circular and Spiral Sweeps.- Occasionally there is some advantage in using a circular or spiral sweep, rather than a linear sweep. A circular sweep may be obtained by applying to the two sets of deflecting plates sine-wave voltages of equal frequency and 90-degree phase displacement. These voltages may be obtained readily by the use of a series combination of resistance and capacitance, as shown in Fig. 5-6 a. The wave under observation may then be superimposed upon the circular time base by producing a radial deflection proportional to the instantaneous voltage. The radial deflection may be produced by varying the second anode voltage and thus the deflection sensitivity. A better method is to amplitude-modulate the original input voltage to the phase-splitting circuit of Fig. 5-6a by the voltage under observation, using a linear modulator.

Fig. 5-6b shows circular sweep patterns which, however, were not obtained by radial deflection but by applying the voltage under observation to the Y plates.



(a)

FIG. 5-6a. Circuit for obtaining circular sweep.



(1) (2) (3)

FIG. 5-6b. Circular sweep patterns: (1) circular sweep alone, (2) wave of frequency 5 times that of sweep voltage superimposed on sweep, (3) wave of frequency 9/2 that of standard. (General Radio Co.)

A spiral sweep is derived from a circular sweep by producing a linear saw-tooth variation of the voltage applied to the phase-splitting circuit. Modulation in accordance with the voltage under observation is superimposed upon this linear increase of sweep voltage amplitude. One of practical difficulties encountered in the design of a spiral sweep is that of causing the linear increase of sweep voltage amplitude to start at exactly the same point in the circular sweep in order that each spiral will coincide with the previous one.

A modified form of the circuit for circular sweep, shown in Fig. 5-7, is useful in the comparison of frequency. The circuit is such that by adjustment of the resistances and capacitances a circular pattern is obtained when each voltage is applied separately. When the two voltages are then applied simultaneously, a pattern is obtained that has one or more loops or cusps. The pattern is stationary if the ratio of the higher to the lower frequency f_2/f_1, reduced to its simplest form, is N_2/N_1, in which N_2 and N_1 are integers. For the circuit of Fig. 5-7 there are $N_2 + N_1$ loops or cusps, pointed outward. In generating the pattern, the spot moves from one cusp or loop to the N_1th next one. Thus N_1 may be determined by adding one to the number of cusps skipped by the spot, and N_2 may be determined by subtracting N_1 from the total number of cusps. If the position of one condenser and its associated resistor is interchanged, the pattern has $N_2 - N_1$ cusps or loops, pointed inward. This type of pattern is usually more difficult to interpret. If the frequency ratio is not rational, the pattern

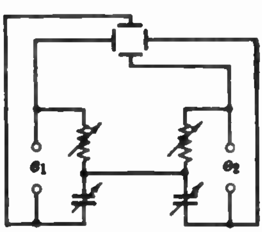


FIG. 5-7. Circuit for the oscillographic comparison of frequency.

rotates and the frequency of rotation must be used in determining the exact frequency ratio.[5]

5-8. Linear Electromagnetic Deflection.- In producing a linear sweep with electromagnetic deflection it is necessary that the current through the coil, and not the voltage impressed across it, shall be of saw-tooth form. Because the voltage across an inductance is proportional to the rate of change of current, the voltage necessary to produce a saw-tooth wave of current through the deflecting coils is not of saw-tooth form.

In order to make it unnecessary for the sweep oscillator to supply the high current required to produce the sweep deflection, the output of the oscillator is impressed upon the grid of a current amplifier, the plate current of which excites the coils, as shown in Fig. 5-8a. The equivalent plate circuit is shown in Fig. 5-8b. Curves A, B, C, and D of Fig. 5-9 show the wave form of the alternating plate current; the voltage across

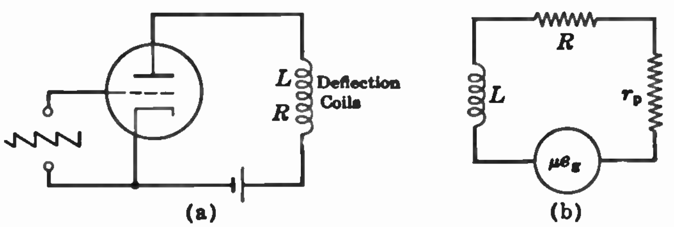


Fig. 5-8. (a) Amplifier for electromagnetic sweep deflection. (b) Equivalent plate circuit.

the inductance, which is proportional to the rate of change of current; the wave form of the voltage across the plate and coil resistance, which is proportional to the current; and the wave form of the ficticious voltage μe_g, which is the sum of the voltage across the inductance and that across the resistance. Since μ is assumed constant, the exciting grid voltage e_g must be of the same form as μe_g. A voltage of the form of Fig. 5-9d can be obtained from a saw-tooth oscillator when a resistance R' is used in series with the condenser, as shown in Fig. 5-10. If the condenser charging current is essentially constant, as required in order that the condenser voltage shall rise linearly, a constant positive voltage appears across the series resistance when the condenser charges, and a larger negative voltage when the condenser discharges. The resistance thus supplies

5. Reich, H.J., "Theory and Applications of Electron Tubes," McGraw-Hill, Sec. 15-21.

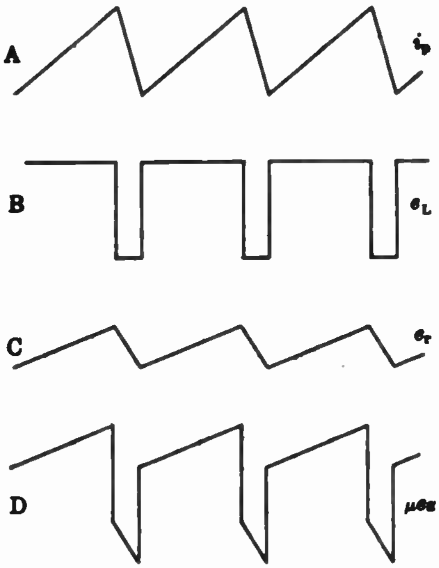


Fig. 5-9. Wave forms of current and voltages in the circuit of Fig. 5-8. (Time of sweep greatly exaggerated.)

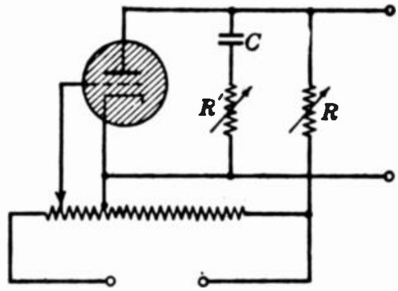


Fig. 5-10. Sweep oscillator for magnetic deflection.

the rectangular component of voltage required in the grid circuit in order to produce a saw-tooth wave of current through the coil. In practice the series resistance R' is varied until the current wave has the desired form.

At sweep frequencies of the order of those normally used in cathode-ray oscillographs the impedance of the deflecting coils is sufficiently low so that appreciable increase of coil current can be obtained by the use of a step-down transformer between the amplifier tube and the coils. This is particularly true when a pentode is used, since the plate resistance of a pentode is high. The variation of load impedance and hence of sensitivity with frequency is one of the disadvantages in the use of magnetic deflection.

The discontinuities in the voltage required to give saw-tooth current result in the generation of transient oscillations in the transformer and deflecting coils. Damped waves of current are therefore superimposed upon the desired saw-tooth wave of current. These can be eliminated by shunting the transformer primary or secondary with a resistance, but this has the objectionable effect of lengthening the return sweep. By shunting the transformer primary with a diode in series with a parallel combination of capacitance and resistance, as shown in Fig. 5-11, the damping is obtained only during the portion of the cycle in which the diode anode is positive. The diode is therefore connected so that it conducts during the forward sweep, when the plate current increases, but not during the return sweep, when the plate current decreases.

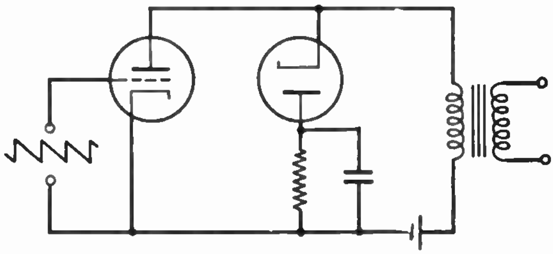


Fig. 5-11. Amplifier for electromagnetic sweep.

5-9. Amplifiers for Cathode-ray Tubes.- Amplifiers are used in connection with the sweep oscillators of cathode-ray oscilloscopes both in order to obtain the high sweep voltages necessary to give adequate deflection when high anode voltage is used and in order to make possible the variation of sweep amplitude without affecting the wave form of the oscillator output. It has been found that an amplifier capable of amplifying a saw-tooth wave without excessive distortion should be able to amplify frequencies ranging from the fundamental up to the tenth harmonic without appreciable frequency or phase distortion. In modern cathode-ray oscillographs the sweep frequency can be varied from a frequency of 50 cycles per second, or less, up to approximately 50 kilocycles. The amplifier must therefore be capable of amplifying without appreciable frequency or phase distortion over a range of from 50 to 500,000 cycles. A similar requirement must be satisfied in amplifying a voltage of the form of Fig. 4-28a used as explained in Sec. 5-6, even though the repetition frequency is comparatively low. In the use of cathode-ray tubes in the study of the wave form of voltages or currents, in television, in ionosphere sounding, and in similar applications, it is equally important that the amplifier used to amplify the Y-deflection voltages shall be capable of amplifying over this wide frequency range without distortion. Amplifiers of this type, which were first designed for the amplification of television signals, are called video amplifiers.[6] Because of the very wide frequency response required of video amplifiers, only direct or resistance-capacitance coupling can be used.

The principal problem encountered in the design of video amplifiers is the loss of amplification at high frequencies relative to midband frequency amplification as the result of tube and circuit capacitances, particularly input capacitance. This problem may be solved in a number of ways. The simplest is the

6. Discussed in Ch. 3.

use of low values of plate and grid coupling resistances, so that the capacitive reactance shunting the coupling resistances is large in comparison with the resistances at the highest frequency to be amplified, and therefore has comparatively little effect. This method achieves uniformity of amplification at the expense of amplification at midband and low frequencies. A second, more satisfactory method consists of the use of a small amount of inductance in series with the plate coupling resistance. The inductance resonates with the tube capacitance at high frequency and thus prevents the amplification from falling. In fact, the amplification may be made to rise at high frequencies by this method. Above resonance the amplification falls rapidly. The range of uniform amplification may be extended to even higher frequency by shunting the inductance by a small capacitance. A third method of obtaining uniform amplification over the required wide frequency band is by the use of a band-pass filter of more complicated form as the coupling network. The fourth method is the use of "cathode-follower" or "unity-amplification" stages in the input of the amplifier and between high-gain stages.

The cathode-follower amplifier, shown in basic form in Fig. 5-12, is a single-stage inverse feed-back amplifier in which the output voltage is taken from across the cathode resistor. By means of the equivalent plate circuit the student can readily show that the voltage amplification of this amplifier is:

$$\underline{A} = \frac{\mu R_k}{r_p + R_k(1 + \mu)} \tag{5-3}$$

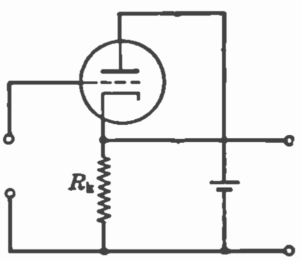


Fig. 5-12. Cathode-follower amplifier.

Equation (5-3) shows that the amplification of a cathode-follower amplifier is always less than unity. An analysis based upon the equivalent plate circuit also shows that the effective input capacitance of such an amplifier is less than $C_{gk} + C_{gp}$, as compared with approximately $C_{gk} + \mu C_{gp}$ in an amplifier in which the output is taken from a resistor in series with the plate. The effective input conductance is also less than that of the ordinary type of amplifier. The output terminal admittance of a cathode-follower stage is

$$\underline{Y}_o = 1/R_k + (\mu + 1)/r_p + j\omega(C_{gk} + C_{pk}) \tag{5-4}$$

Since the second term of eq. (5-4) is of the order of 1000 ohms or less, the third term is negligible by comparison at frequencies up to about 2 megacycles. Between the output terminals, therefore, the amplifier acts like a pure resistance of value less than $R_k/(1 + R_k g_m)$. Since most modern tubes have a transconductance in excess of 1300 micromhos, the effective output resistance may be readily made less than 750 ohms, and be as low as 170 ohms. Advantages attainable as the result of the low input capacitance, high input impedance, and low output terminal impedance, far offset the loss of amplification in the use of cathode-follower stages in video amplifiers. Because the cathode-follower amplifier is a degenerative amplifier, it introduces negligible amplitude distortion and is capable of handling a high input voltage without overloading.

Figure 5-13 shows a typical amplifier incorporating cathode-follower stages. Use of the cathode-follower circuit in the input insures that low current and power will be drawn from the source of the voltage under observation, and minimizes the danger of distortion as the result of current drain or phase

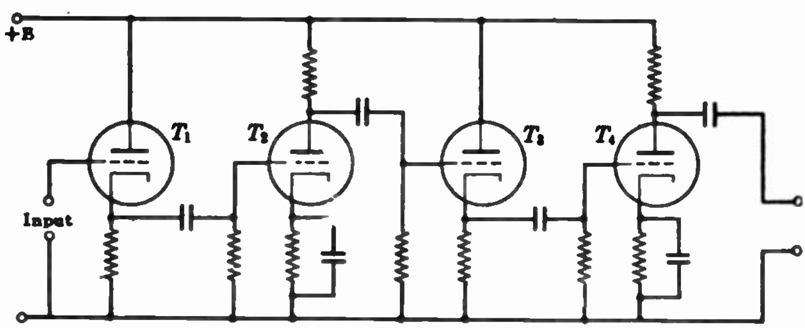


FIG. 5-13. Resistance-capacitance-coupled amplifiers incorporating cathode-follower input and coupling (T_3) stages.

shift caused by input capacitance. Furthermore, because of the low output impedance of this stage, the relatively high effective input capacitance of the second, high-gain, stage has no effect upon the amplification and phase shift at frequencies below 500 kilocycles. The use of the cathode-follower third stage results in low effective capacitance in shunt with the grid and plate coupling resistors of the second stage. With the aid of the coupling inductance the amplification may be made constant and the phase shift small up to the required high-frequency limit of amplification. The low output terminal impedance of the third stage also prevents phase and frequency distortion as the result of the input capacitance of the final stage.

It is of interest to note that the cathode-follower amplifier is also very useful in coupling to a load having a low impedance, such as a transmission line. The advantage of the amplifier over a transformer lies in the fact that output terminal impedance of the cathode-follower amplifier is independent of frequency.

In order to prevent slowing up of the electron beam, at least one of each set of deflecting plates must be connected to a positive point of the cathode-ray tube voltage supply, usually to the second anode. Figure 5-14 shows one circuit used in coupling the output of an amplifier to the deflecting plates. By means of the adjustable voltage divider the direct voltage between the deflecting plates can be changed, and the pattern thus be displaced on the screen. Because the two deflecting plates are not at the same potential relative to the anode in this circuit, the deflecting field is distorted in a manner that causes defocusing of the spot when it is deflected. For this reason it is now the usual practice to use a push-pull amplifier in applying the excitation to the deflecting plates. Figure 5-15 shows a circuit that can be used for this purpose. This circuit incorporates cathode phase inversion.[7] The alternating voltage

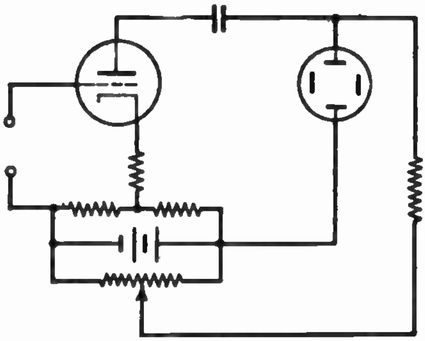


FIG. 5-14. Oscillograph amplifier with control for displacing pattern.

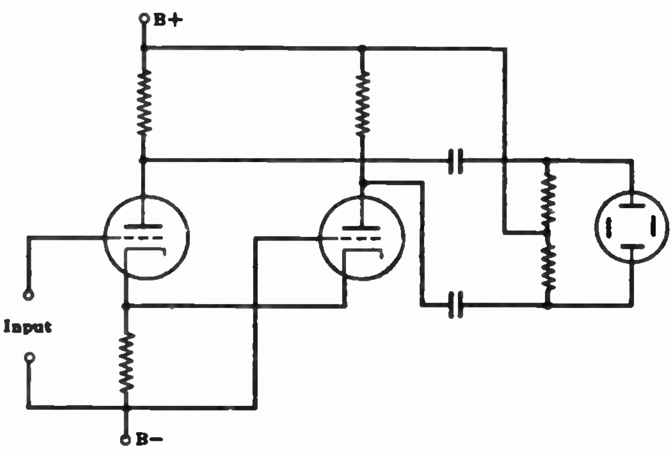


FIG. 5-15. Use of cathode phase-inverter to excite cathode-ray tube.

7. Schmitt, O.H., J. Sci. Inst. 15, 136 (1938); Rev. Sci. Instr., 12, 548 (1941)

produced across the cathode resistor as the result of the grid excitation of V_1 is impressed upon the grid of V_2 in such polarity as to make the grid excitation of V_2 180 degrees out of phase with that of V_1. The output voltages of the two tubes are therefore opposite in phase. By proper choice of resistances the output voltages can be made equal in amplitude. Fig. 5-16 shows a single-sided transformer-coupled circuit that will operate satisfactorily over a narrower band of frequencies.

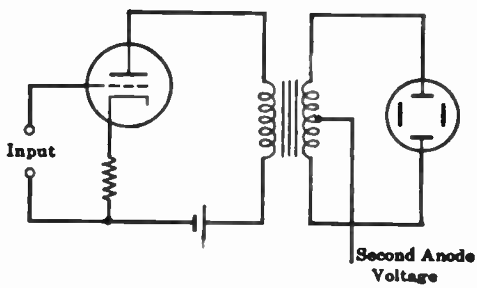


Fig. 5-16. Transformer coupling to cathode-ray tube.

Figure 5-17 shows the circuit diagram of a modern cathode-ray oscillograph. The student should note the following features:

1. The use of cathode-follower circuits in the first and fourth stages of the Y-deflection (vertical) amplifier. A single twin-triode tube is used for these two stages. The first and second stages of the sweep amplifier also use the cathode follower circuit. The output of V_8 is applied through a pulse-sharpening circuit, C29 and R49, to the grid of the cathode-ray tube in order to eliminate the return sweep. Because of the low output terminal impedance of the cathode follower stage V_8, the pulse-sharpening condenser does not cut down on the high-frequency amplification. If an amplifier stage of the ordinary type were used, the voltage pulse would be applied to the grid of the cathode-ray tube in the wrong polarity. The functions of the cathode follower stages V_1, V_4, and V_9 have been discussed earlier in this section.

2. The use of inductances L_1 to L_6 in the coupling circuits in order to improve the frequency response.

3. The use of cathode phase inversion in the push-pull amplifiers V_5-V_6 and V_{10}-V_{11}.

4. The manner in which the deflecting electrodes are connected through high resistances to a positive point in the cathode-ray-tube power supply.

5. The use of a condenser-resistance filter C_{31}-R_{53}-C_{30} in the output of the half-wave rectifier V_{13} that supplies voltages to the cathode-ray tube. A similar filter C_{35}-R_{56}-C_{36} is used in the output of the half-wave rectifier V_{15} that supplies biasing voltage to the amplifiers V_1 and V_4.

6. The use of a single secondary winding to supply the three rectifiers V_{13}, V_{14}, and V_{15}.

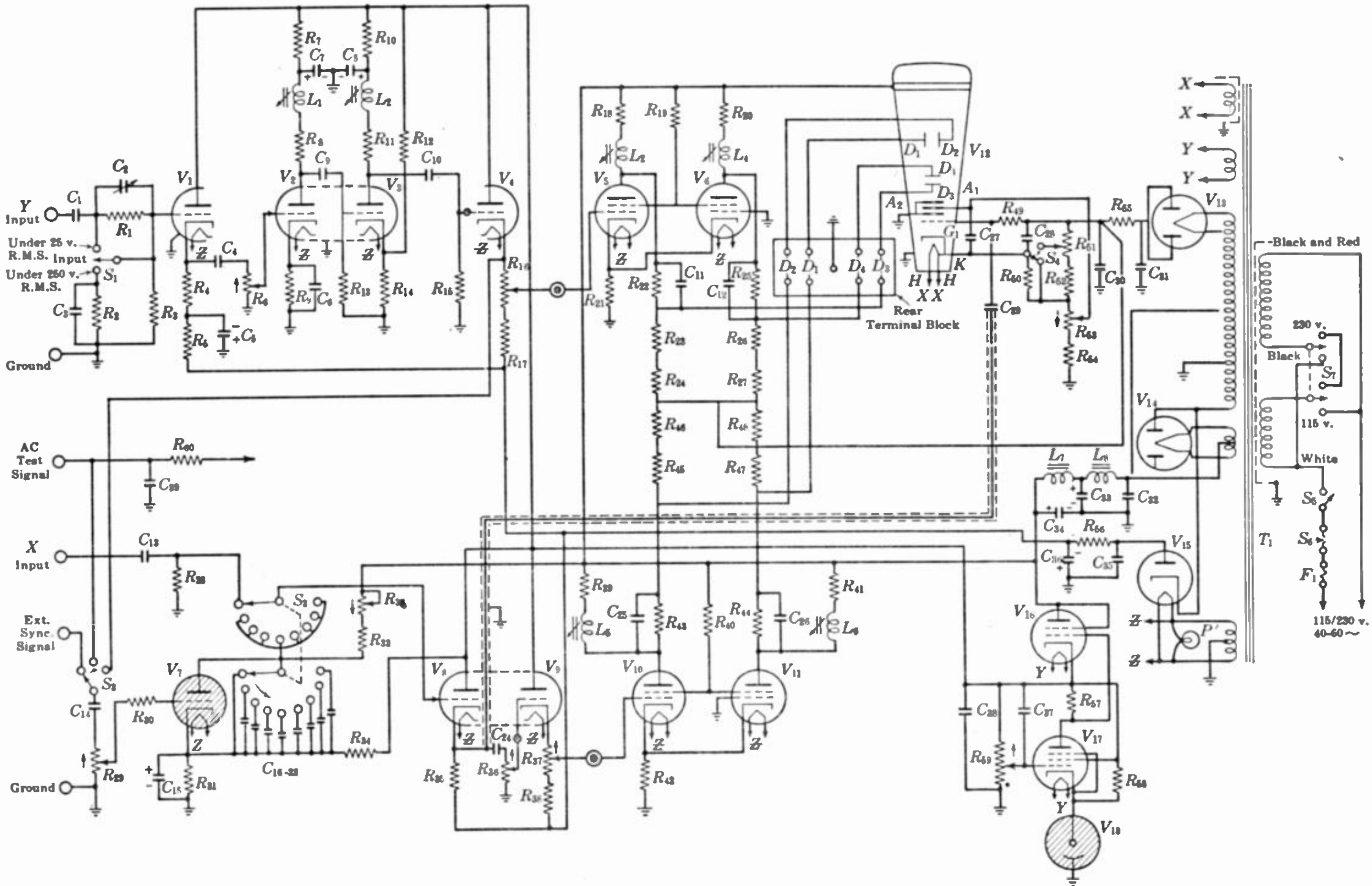


FIG. 5-17. Dumont cathode-ray oscillograph. (Type 208).

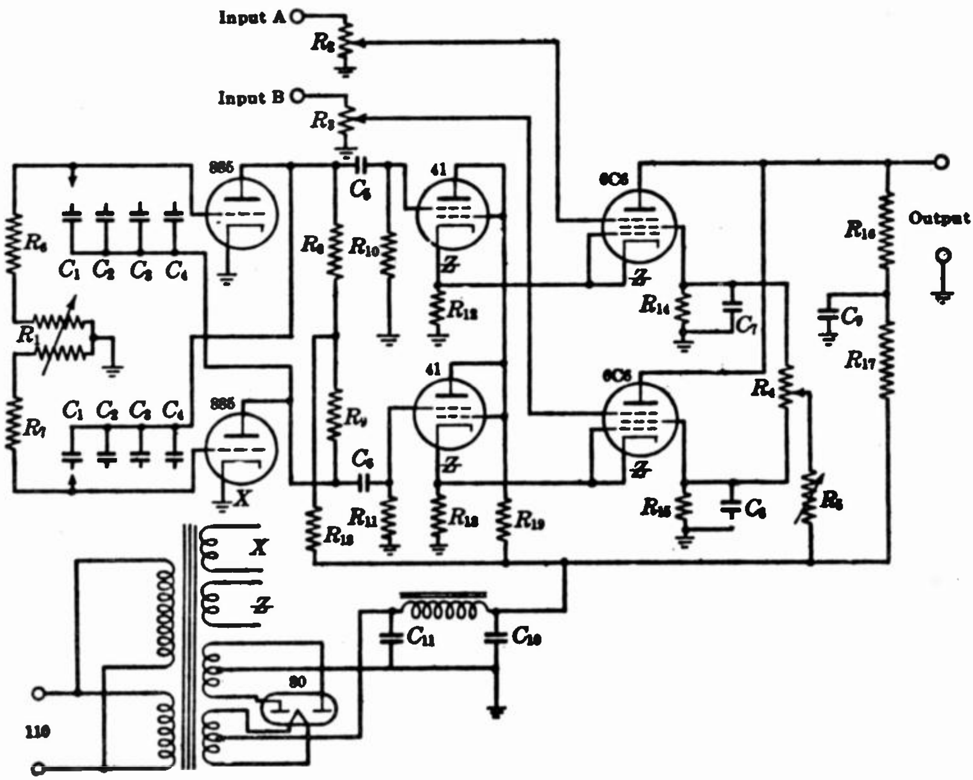


FIG. 5-18. Dumont type 185 electron switch and square-wave generator.

7. The voltage stabilizer V_{16}-V_{17}, which is of the general form of that of Fig. 2-13.

8. The switch S_4 used to apply a high negative bias to the grid of the cathode-ray tube in order to turn off the beam without changing the adjustment of the intensity control R_{51}.

5-10. Electronic Switches.- A useful device that is often used in conjunction with cathode-ray oscillographs is the electronic switch, which makes possible the simultaneous observation of two waves on a single screen. The device consists essentially of two amplifiers the outputs of which are connected in parallel to the plates of the cathode-ray tube. The two voltages to be compared are impressed upon the two amplifiers, which are biased alternately to cut off by means of square-wave biasing voltage impressed in opposite phase in one of the grid circuits of each of the amplifiers. If the square wave is symmetrical

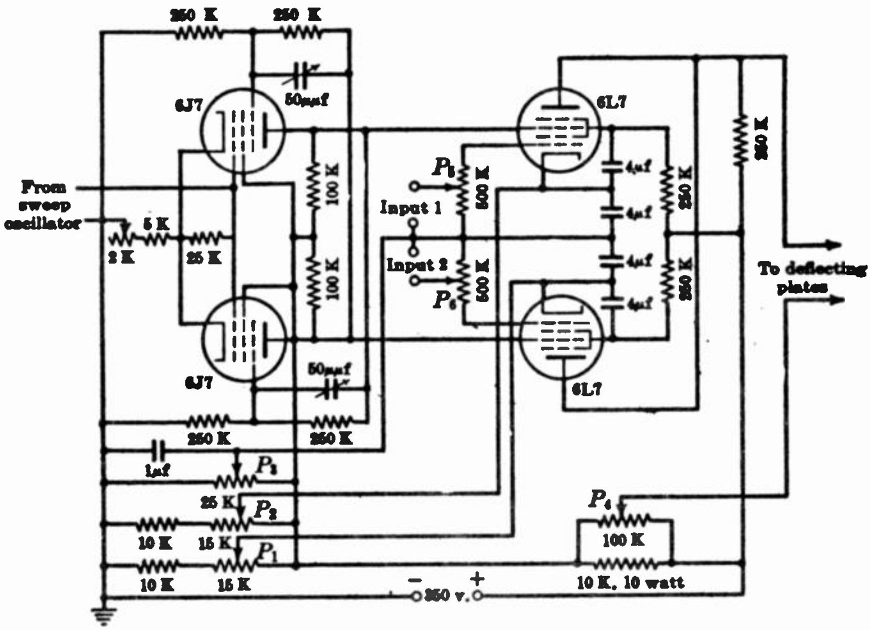


FIG. 5-19. Circuit diagram of electronic switch. "K" represents one thousand ohms.

each amplifier amplifies during half the time. The square-wave generator may be synchronized to the sweep oscillator in such a manner that each amplifier operates during alternate sweeps, or the switching may be accomplished at a higher or lower frequency. If the waves under observation are of low frequency, objectionable flicker can be avoided by switching at a frequency many times the sweep frequency. The two patterns then appear on the screen as dashed or dotted lines. Figure 5-18 shows the circuit of an electronic switch that is driven by a separate source of switching voltage. The amplitudes of the individual waves are adjusted by means of P_1 and P_2. The patterns are displaced relative to each other by means of P_3. Figure 5-19 shows a circuit that may be driven by means of the sweep oscillator, or by a separate saw-tooth or square-wave oscillator coupled through a pulse-sharpening circuit.[8]

8. Reich, H.J., Rev. Sci. Instr., 12, 191 (1941). With bibliography of 9 items.

BIBLIOGRAPHY ON CATHODE-RAY OSCILLOSCOPES

Rider, J.F., "The Cathode-ray Tube at Work," J.F. Rider, New York, 1935

Maloff, I.G., and Epstein, D.W., "Electron Optics in Television," McGraw-Hill Book Company, Inc., New York, 1938

Watson-Watt, R.A., Herd, J.F., and Bainbridge-Bell, L.H., "The Cathode-ray Oscillograph in Radio Research," H.M. Stationery Office, London, 1933

Ardenne, M. von, "Die Kathodenstrahlrohren," Julius Springer, Berlin, 1934

MacGregor-Morris, J.T., and Henley, J.A., "Cathode Ray Oscillography," Chapman and Hall, Ltd., London, 1936

Knoll, M., Arch. tech. Mess., 1, 76 (1931) (summary of literature to 1931)

Zworykin, V.K., Electronics, November, 1931, p. 188 (bibliography of 12 items)

Batcher, R.R., Proc. I.R.E., 20, 1878 (1932) (bibliography of 29 items)

Stinchfield, J.M., Elec. Eng., 53, 1608 (1934)

Batcher, R.R., Instruments, 9, 6, 38, 77, 112, 140, 166, 197, 231, 255, 286, 312, 341 (1936)

Parr, G., "The Low Voltage Cathode Ray Tube," Chapman and Hall, Ltd., London, 1937 (extensive bibliography)

Stocker, A.C., Proc. I.R.E., 25, 1012 (1937)

Mayer, H.F., Electronics, April, 1938, p. 14

Preisman, A., RCA Rev., 3, 473 (1939)

Overbeck, W.P., and Löf, J.L.C., Rev. Sci. Instruments, 11, 375 (1940)

Geohagen, W.A., Electronics, November, 1940, p. 36

Chapter 6

MODULATION

The effective transmission by radio means of a signal which originally is picked up by a microphone in telephony, or is created by the operation of a key or switch in telegraphy, or is generated by a separate electronic device in the case of television or the "terrain clearance meter" for determining the height of a plane above the earth, or is otherwise introduced into an electric circuit, can only be accomplished if the frequency is considerably above the audio range. The electric signal generated in a microphone from a sound wave may be transmitted directly to the receiving station over a wire transmission line. On the other hand, this same signal cannot be transmitted to the receiving station by radiation through space with an efficiency sufficiently high to be feasible and without a prohibitive amount of interference if more than one transmission takes place at the same time. Consequently, the average frequency of the signal must be raised, usually by a large amount. It is shown in Ch. 12 (Radiation) that the energy radiated for a given amount of input power to an antenna goes up as some positive power of the frequency, so that the energy that can be radiated at audio frequencies is practically negligible, whereas the energy that can be radiated at 1 megacycle (1,000,000 cps; this is in the commercial broadcast band) for the same input power is quite appreciable. In order to radiate a plurality of signals simultaneously without having one interfere with the other at the receiver, the signal sent out by one transmitter must have a frequency range not overlapping that of any other transmitter the radiation from which can also be received by the receiver.

A change of the average frequency of the input signal is not only desirable in radio communication but also is frequently useful in other types of communication. In telephony or other signaling using conductors, numerous signals may be transmitted separately over a single suitably designed two-conductor circuit provided that in transmission the various signals are restricted to frequency ranges that do not overlap. Such a system is called carrier telephony, and in some two-conductor circuits as many as 200 separate signal channels are available. Suitable filters must be used at the receiving end of the system to separate the different channels.

Again, in a superheterodyne receiver the average frequency of an incoming signal, whatever it may be, is reduced to a fixed value to utilize a fixed-frequency intermediate amplifier. This process is closely allied to the process by which frequency changes requisite for radiation, carrier transmission, etc., are brought about.

6-1. Definition and types of modulation.- Modulation is the process of producing a wave some characteristic of which varies as a function of the instantaneous value of another wave, called the modulating wave.[1] The term "wave" is used here as in sine-wave current or voltage, and is a generic term indicating current, voltage, or other electrical quantity varying with time. The modulating wave is usually the signal; the resultant of modulation is the modulated wave which is the wave having some characteristic "which varies as a function of the instantaneous value"[1] of the modulating wave.

Consider a wave (current, or voltage, or other time-varying electrical quantity).

$$A \cos (\omega t + \theta) \qquad (6\text{-}1)$$

in which t is time. If either A, θ, or ω, or any function such as a derivative or integral of one of these, is varied in accordance with some function of the instantaneous value of some other wave (the modulating wave), then (6-1) represents a modulated wave. The types of modulation possible are thus many, but at the present time only two or three are practically important. These are amplitude modulation (A-M), in which A of (6-1) is varied in accordance with the modulating wave, while ω and θ remain constant; frequency modulation (F-M), in which the frequency is varied while A and θ remain constant; and phase modulation (only θ varied) which is usually considered in connection with F-M. The essential characteristics of the various types of modulation will be set out below, after which the means of achieving them will be discussed.

6-2. Characteristics of Amplitude Modulation.- Let $e_s = E_s \cos \omega_s t$ be a signal voltage, and let $e_o = E_o \cos (\omega_o t + \theta)$ be a voltage independent of e_s, called the carrier voltage; $\omega_o/2\pi$ is the carrier frequency, f_o, assumed to be much greater than the signal frequency $f_s = \omega_s/2\pi$, and to be in the range suited to the problem at hand; that is, to the channel available or the frequency range in which it is desired to transmit. Then

$$e_a = E_a (1 + kE_s \cos \omega_s t) \cos (\omega_o t + \theta) \qquad (6\text{-}2)$$

is an amplitude-modulated wave, in which E_o of e_o has been

1. Standards on Transmitters and Antennas, 1938, I.R.E., p. 3.

replaced by $E_a(1 + kE_s \cos \omega_s t)$ where E_a and k are constants dependent on E_o and the electrical system producing the modulation. It is customary to write the amplitude-modulated wave in the form

$$e_a = E_a(1 + m_a \cos \omega_s t) \cos (\omega_o t + \theta) \tag{6-3}$$

where m_a is the modulation factor, and 100 m_a is the percentage modulation.

Expanding the expression (6-3) for e_a gives

$$e_a = E_a \cos (\omega_o t + \theta) + \frac{1}{2} E_a m_a \cos [(\omega_o + \omega_s)t + \theta] + \frac{1}{2} E_a m_a \cos [(\omega_o - \omega_s)t + \theta] \tag{6-4}$$

in which the first term is of carrier frequency, the second has a frequency called the upper side-frequency, equal to the sum of the carrier and signal frequencies, and the last has a frequency called the lower side-frequency equal to the carrier minus the signal frequency. Thus e_a given by (6-3) is the sum of three waves each of different frequency.[2] Illustrations of amplitude-modulated waves discussed here are pictured in Fig. 6-1. The signal e_s is contained in (6-3), but the equivalent three waves of (6-4) show that the frequencies of the resultant simple sine waves are in the neighborhood of ω_o rather than ω_s.

It is worth noting that (6-3) is not necessarily the equivalent of a Fourier analysis. On the other hand, the coefficient $E_a(1 + m_a \cos \omega_s t)$ of the carrier wave in (6-3) is periodic, and in the practical case in which $\omega_o \gg \omega_s$, the coefficient is the envelope of the modulated wave, as indicated in Fig. 6-1. The signal may thus be considered to be contained in the varying amplitude of $\cos (\omega_o t + \theta)$ as indicated by (6-3), i.e. in the envelope of the modulated wave as indicated by Fig. 6-1, or alternatively in the carrier[3] and the two side-frequency sine waves, as indicated in (6-4). All these are equivalent.

If $\omega_o/2\pi$ is made the frequency desired for transmission, the signal e_s has been combined with the carrier to give a resultant (6-4) in which the waves are simple sine waves with frequencies at or near the frequency of transmission. If (6-4) is to be radiated or otherwise transmitted, a channel $2\omega_s/2\pi$ wide must be provided,[4] and suitable apparatus at the receiving end

2. When the signal contains numerous terms of different frequencies, the group of terms of frequencies of the type $\omega_o + \omega_s$ is called the upper side band, that of type $\omega_o - \omega_s$ the lower side band.

3. Note that the carrier is not fundamentally essential to communication since the signal is contained in the two side bands alone.

4. If the signal contains terms of different frequencies, ω_s must be taken as the greatest value of ω_s to be transmitted.

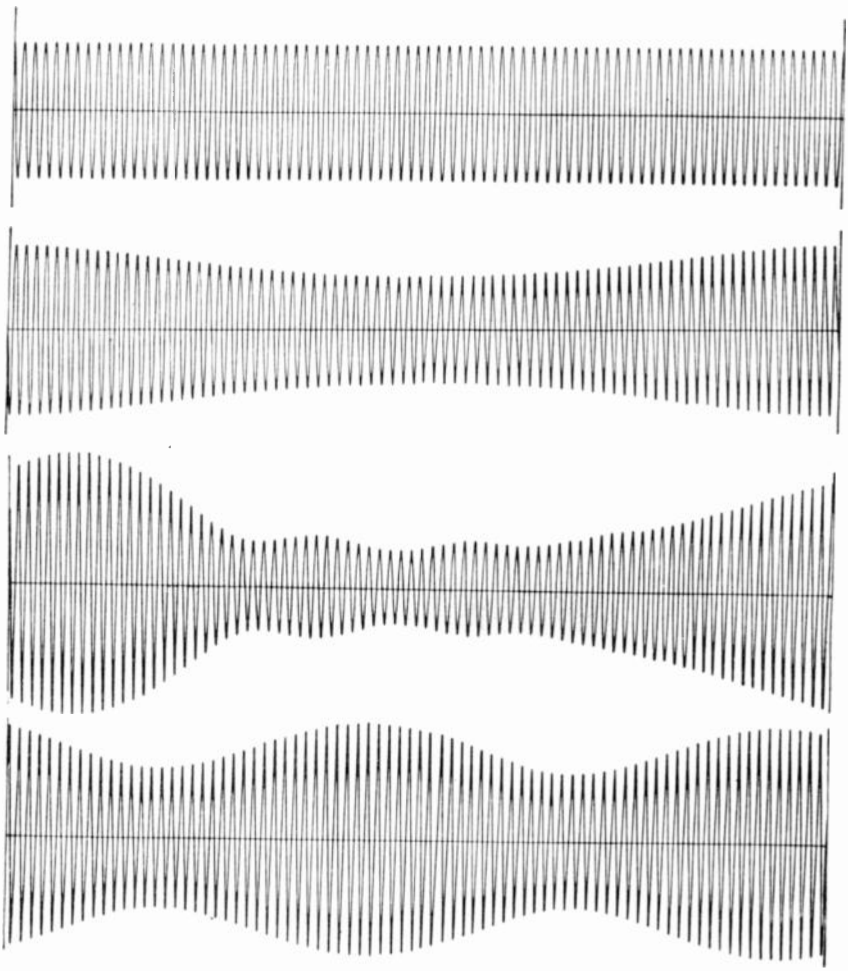


FIG. 6-1. Amplitude-modulated waves. Curve (a) is a carrier, curves (b), (c), and (d) are amplitude-moduated waves. Each modulating wave (not shown) has a form identical with that of the envelope of the corresponding modulated wave. For curves (b) and (d) the modulation is approximately 25%; for curve (c) 60%. (Note: the carrier for curve (c) is of the same frequency as (a) but of a greater amplitude.)

must be supplied to change a received wave proportional to e_a back to one proportional to e_s (Ch. 7).

If the instantaneous voltage e_a were applied across a fixed resistance R, the power dissipated would be $e_a{}^2/R$, and the average power dissipated would approximate $E_a{}^2/2R$ corresponding to the first term on the right-hand side of (6-4), plus $m_a^2 E_a^2/4R$ corresponding to the two side-frequency terms. Since the amplitude of the signal E_s is contained in m_a, and ω_s in the side-frequency terms, the average power corresponding to the signal would be greatest, in relation to the average carrier power, when m_a was made as large as possible. The maximum for m_a is unity (100 per cent modulation), and the above argument shows in a rough way that for a fixed carrier intensity, the greatest amount of energy-bearing information concerning the original modulating signal will be received when modulation is 100 per cent. If the original signal, as in the transmission of speech, contains many terms and is not a pure sine wave, not all components of the signal can be 100 per cent modulated, but the maximum modulation should be 100 per cent. Figure 6-1c shows a modulated wave in which the modulating signal is not a pure sine wave.

6-3. Characteristics of Frequency Modulation.- If, in (6-1), ω were made to vary directly as $e_s = E_s \cos \omega_s t$, the product ωt would depend on t, so that different results would be obtained at different times with the same signal. This is undesirable and is not used. However, when the frequency is a constant, the complete argument φ of the cosine function $A \cos \varphi$ may be considered as resulting from the operation $\varphi = \int \omega\, dt$, and this gives

$$A \cos \left(\int \omega\, dt\right) = A \cos (\omega t + \theta) \quad (\omega = \text{const.}) \qquad (6\text{-}5)$$

where θ is the constant of integration. By extending this notion to include cases where the frequency is varied, one obtains a useful[5] concept of frequency, namely, the <u>instantaneous frequency</u> of a sine wave with a varying frequency is the ω under the integral sign in (6-5), or, written in a slightly different form:

$$\omega = d\varphi/dt \qquad (6\text{-}6)$$

Consider a signal voltage $e_s = E_s \cos \omega_s t$ and a carrier

5. But approximate; the concept of frequency as the derivative of the argument of the sine wave is useful here but breaks down in certain other cases, and particularly at low frequencies. Use of a constant frequency is so common that a word of explanation concerning variable frequency may be desirable. The definition of "frequency" when frequency is varying is not easy. A simple definition is that instantaneous frequency is 2/T where T is the time between two adjacent zeros. This definition has serious drawbacks. In seeking for another definition which, when ω is constant will give the usual results, (6-5) and (6-6) appear logical and easy to handle. But such a definition does not assure that common ideas of frequency can be applied to the varying frequency case, and leads to such anomalies as that discussed in the next footnote.

voltage $e_o = E_o \cos(\omega_o t + \theta)$. If, now, by an appropriate circuit ω is made to vary so that

$$\omega = \omega_o(1 + kE_s \cos \omega_s t) \tag{6-7}$$

$$= \omega_o(1 + m_f \cos \omega_s t) \tag{6-8}$$

then the wave $$e \doteq E \cos \left(\int \omega\, dt\right) \tag{6-9}$$

is a modulated wave, called a frequency-modulated wave, in which the signal voltage e_s is contained in the angular frequency ω given by (6-7). In (6-7) and (6-8), ω_o is the constant carrier frequency (in commercial F-M broadcasting ω_o is of the order of magnitude of 45 Mc) and $m_f\omega_o/2\pi = m_f f_o$ is the frequency deviation.[6] It is essential to keep in mind that m_f depends on E_s but not on ω_s.

Although the idea of defining an instantaneous frequency as above is useful, the actual application of a voltage of the form of (6-9) to a circuit, or the passage of a current of this form in the circuit, or the radiation of fields due to a current of this form, is most easily handled by breaking (6-9) into simple sinusoidal waves of constant frequency. This is analogous to the procedure in the A-M case, where the wave (6-3) of varying amplitude was split into three waves [carrier and upper and lower side-frequency waves, each having constant amplitude, frequency, and phase angle; see eq. (6-4)].

From (6-9) and (6-8)

$$e = E \cos \left(\int \omega\, dt\right) = E \cos \left[\omega_o t + \omega_o \frac{m_f}{\omega_s} \sin \omega_s t + \theta\right] \tag{6-10}$$

Write the last expression

$$e = E \cos (\omega_o t + \theta + a \sin \omega_s t)$$
$$= E[\cos (a \sin \omega_s t) \cos (\omega_o t + \theta) - \sin (a \sin \omega_s t) \sin(\omega_o t + \theta)] \tag{6-11}$$

where $a = \omega_o m_f/\omega_s$. There are expansions for quantities of the type $\cos (a \sin \omega_s t)$ and $\sin (a \sin \omega_s t)$, which, although not particularly common, nevertheless are not extremely difficult to deduce.[7] These are

6. Although $m_f f_o$ is called the frequency deviation, it is not the greatest difference between the carrier frequency and the side frequencies discussed below. The frequency deviation is the maximum change in the instantaneous frequency defined in (6-6), but this does not say that a group of sine-wave components into which the F-M wave can be resolved will be limited to frequencies within $m_f f_o$ of the carrier.

7. Gray, Mathews, and MacRobert, "Bessel Functions," Macmillan and Co., 1931, p. 32.

$$\frac{1}{2}\cos\,(a\sin\omega_s t) = \frac{1}{2}J_0 + J_2\cos 2\omega_s t + J_4\cos 4\omega_s t + \ldots \qquad (6\text{-}12)$$

and
$$\frac{1}{2}\sin\,(a\sin\omega_s t) = J_1\sin\omega_s t + J_3\sin 3\omega_s t + \ldots \qquad (6\text{-}13)$$

where each J is a constant which, however, changes with change in a; that is, each J is a quantity depending on a only.[8]

Substituting (6-12) and (6-13) into (6-11) and combining terms trigonometrically, there results

	Frequencies of terms
$\frac{e}{E} = J_0\cos(\omega_0 t + \theta)$	ω_0 (carrier)
$+ J_1\cos[(\omega_0 + \omega_s)t + \theta] - J_1\cos[(\omega_0 - \omega_s)t + \theta]$	$\omega_0 + \omega_s$ and $\omega_0 - \omega_s$
$+ J_2\cos[(\omega_0 + 2\omega_s)t + \theta] + J_2\cos[(\omega_0 - 2\omega_s)t + \theta]$	$\omega_0 + 2\omega_s$ and $\omega_0 - 2\omega_s$
+ etc. (6-14)	$\omega_0 + 3\omega_s$, etc. and $\omega_0 - 3\omega_s$, etc.

Thus e may be resolved into terms of carrier frequency ω_0, upper side frequencies $\omega_0 + \omega_s$, $\omega_0 + 2\omega_s$, etc., and lower side frequencies $\omega_0 - \omega_s$, $\omega_0 - 2\omega_s$, etc. It should be noticed that the negative sign before the term of frequency $\omega_0 - \omega_s$ indicates that this side frequency has the opposite phase from that of its mate of frequency $\omega_0 + \omega_s$. The F-M wave is similar to the A-M wave in that it has a carrier and upper and lower side bands, but differs in the relative phases of the side frequencies and in their number, which for a sine-wave signal is many more than two. The number of side frequencies is determined by the fact that calculation of the J's shows that the J's begin to approach zero in value when n is larger than a, so that the large number of side frequencies indicated in (6-14) is limited by this fact. But almost always F-M requires more side frequencies than the two of A-M, and for this and other reasons F-M channels in commercial broadcasting are 200 kc wide, compared with the 10-kc width of A-M channels. Figure 6-2 shows the relative size of the various sine-wave components of several F-M waves. Pieracci[9] gives pictures obtained with the use of an F-M monitor which shows the amplitude and frequency distribution on a cathode-ray oscilloscope. A trace of one such picture is given in Fig. 6-3 and should be compared with (e) of Fig. 6-2.

8. All the J's can be computed from the one formula

$$J_n = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+n)!}\left(\frac{a}{2}\right)^{n+2k}$$

The J's (Bessel Functions) are used again in Ch. 14. Graphs of J_0 and J_1 are given in Fig. 14-8 for values of a between 0 and 10.

9. Pieracci, Roger J., Proc. I.R.E. 28, No. 8, Aug. (1940).

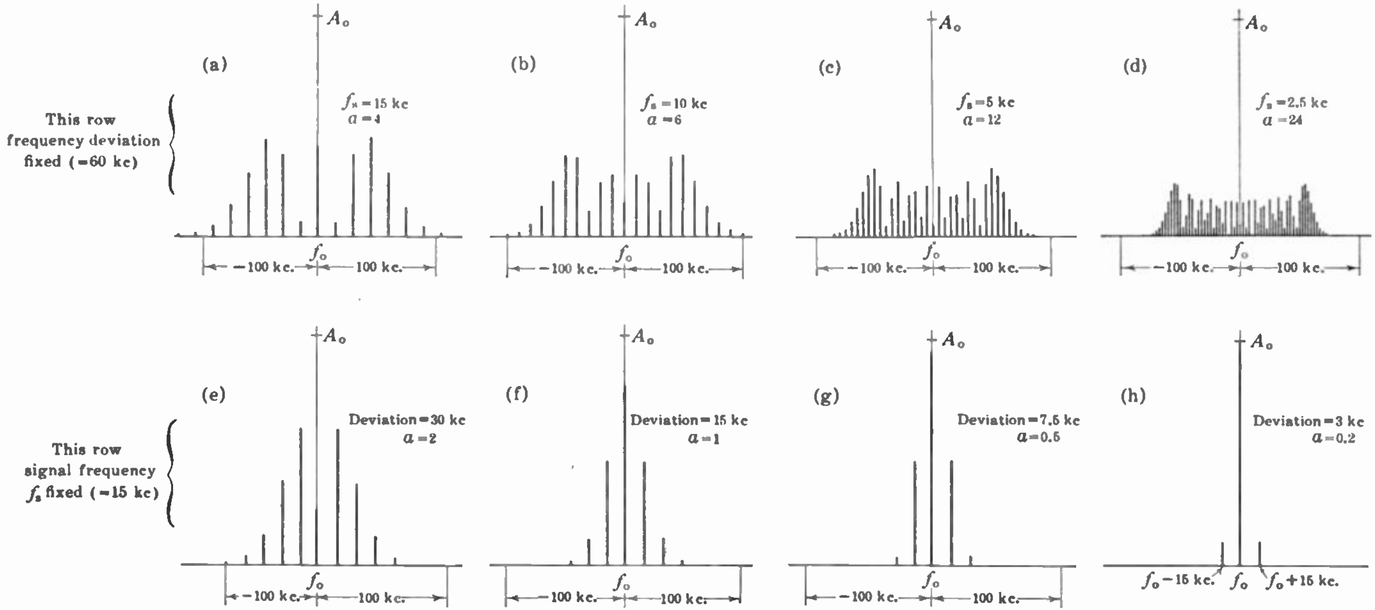


FIG. 6-2. Relative magnitudes of unmodulated carrier, and of the carrier and side-frequency simple sine wave components of a frequency-modulated wave. Figures (a) to (d) are for a frequency deviation ($\omega_o m_f/2\pi$; see eq. 6-9) of 60 kc and various signal frequencies; figures (e) to (h) are for a signal frequency $f_s = \omega_s/2\pi$ of 15 kc and various deviations. A_o in each case indicates the relative amplitude of the *unmodulated* carriers. (From Everitt, Electrical Engineering, Nov. 1940, and Pieracci, Proc. I. R. E., Aug. 1940).

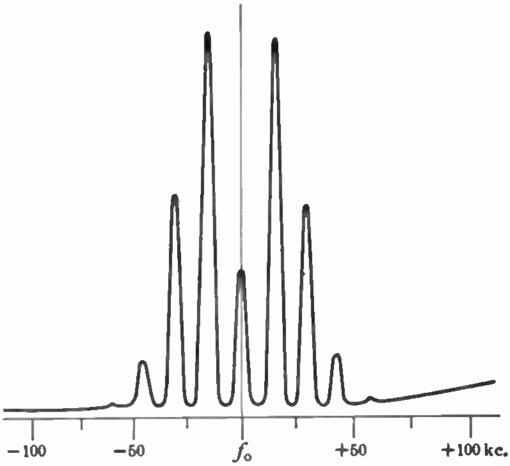


FIG. 6-3. Relative amplitude of the sine-wave components of an F-M wave (eq. 6-11); deviation $(m_f f_o) = 30$ kc, signal frequency $f_s = 15$ kc. Curve is trace of an oscillogram obtained on an F-M monitor. Compare (e) of Fig. 6-2. (From Pieracci, Proc. I. R. E., Aug. 1940.)

6-4. Phase Modulation.- Phase modulation is quite similar to frequency modulation. If in

$$e = E\cos(\omega_o t + \theta) \quad (6\text{-}15)$$

the phase angle θ is varied in direct proportion to the signal $e_s = E_s \cos \omega_s t$ while E and the carrier frequency ω_o remain constant, the signal is effectively contained in the result below, (6-16), and the latter constitutes a phase-modulated wave

$$e = E\cos(\omega_o t + \theta_o + m_p \cos\omega_s t) \quad (6\text{-}16)$$

Here θ_o is a constant and m_p, directly proportional to E_s, is the modulation factor for phase modulation. Comparing (6-16) with (6-10), it is seen that phase modulation differs from F-M only in the fact that m_f/ω_s is the coefficient of the sinusoidal term in (6-10) whereas m_p is the corresponding quantity in (6-14). If by an appropriate circuit a factor m_p/ω_s can be introduced in place of m_p in (6-14), the phase modulation becomes F-M. This procedure is a practical one, carried out in the case of one F-M system.

Because of the close connection between phase modulation and F-M, the former will not be discussed further.[10]

6-5. Types of Amplitude Modulation.- Amplitude modulation is usually produced in a non-linear circuit, although the term linear modulation is used in some cases to denote a direct proportionality between certain factors.

Square-law or Small-signal Modulation. This type of modulation can be illustrated by using the series expansion for plate current given in Ch. 2, namely,

$$i_p = a_1 e + a_2 e^2 + a_3 e^3 + \ldots\ldots \quad (6\text{-}17)$$

where the a's are constants, the value of each depending upon

10. For use of a phase-modulation system in practice, see Crosby, Murray G., "Communication by Phase Modulation," Proc. I.R.E., 27, No. 2, Feb. (1939).

the tube characteristics and external circuits, and $e = e_g + v_p/\mu$ in which e_g is the varying grid-cathode voltage and v_p is the introduced voltage in the plate circuit and μ the amplification factor of the tube.

If, by introducing the signal voltage $e_s = E_s \cos \omega_s t$ and the carrier voltage $E_o \cos (\omega_o t + \theta)$ in either the grid or plate circuit, or one in the grid and the other in the plate circuit, the equivalent voltage e in (6-15) becomes

$$E_s \cos \omega_s t + E_o \cos (\omega_o t + \theta)$$

or a similar expression with $1/\mu$ before one or both terms, then

$$e^2 = \frac{1}{2}E_s^2 + \frac{1}{2}E_s^2 \cos 2\omega_s t + \frac{1}{2}E_o^2 + \frac{1}{2}E_o^2 \cos^2 2(\omega_o t + \theta)$$

$$+ E_s E_o \cos [(\omega_o + \omega_s)t + \theta] + E_s E_o \cos [(\omega_o - \omega_s)t + \theta] \quad (6\text{-}18)$$

Assuming that $\omega_o \gg \omega_s$ and that frequencies not near ω_o can be eliminated by tuned circuits or filters, it is seen that the last two terms in e^2 give the upper and lower side frequencies. Thus the plate current contains numerous components of which those with frequencies near ω_o are

$$i_p = a_1 E_o \cos (\omega_o t + \theta) + a_2 E_o E_s \cos [(\omega_o + \omega_s)t + \theta]$$
$$+ a_2 E_o E_s \cos [(\omega_o - \omega_s)t + \theta] \quad (6\text{-}19)$$

plus possible contributions from $a_3 e^3$, $a_4 e^4$, etc. Equation (6-19) is of the same form as (6-4), hence amplitude modulation has been achieved provided the i_p given by (6-19) can be separated from the total i_p, as is possible.

It is seen that operation of a modulator on the principle outlined above depends on emphasizing the coefficient a_2. When no grid current flows and μ is constant and the load is pure resistance r_o,

$$a_2 = \frac{-\mu^2 r_p r_p'}{2(r_o + r_p)^3} \quad (6\text{-}20)$$

where $r_p' = \partial r_p/\partial e_b$. Thus a_2 is greater the greater the curvature of the $i_b - e_b$ characteristic since $\partial i_b/\partial e_b = 1/r_p$, and $\partial r_p/\partial e_b = -r_p^2(\partial^2 i_b/\partial e_b^2)$. When the load contains reactance, or when μ varies, or i_g is not zero, or other change occurs from the simple case considered above, amplitude modulation is still obtained but a_2 is a rather complicated function of the tube and associated circuit parameters.[11]

The amount of modulated output available without appreciable distortion in a square-law modulator of the type described above is not great, and the efficiency is low. Numerous circuits

11. McIlwain and Brainerd, "High Frequency Alternating Currents," Appendix A.

can be devised: introducing the signal voltage into the grid or plate circuit and the carrier into the grid or plate circuit, having the grid negative or positive, applying one voltage to one grid and the other to another grid of a multielectrode tube, etc. Likewise a non-linear impedance can be used for modulation, since the current through the device can usually be related to the voltage e across it by an equation of the form of (6-17) when e is small. But present-day amplitude modulation is often produced by varying the potential or bias of an electrode of an amplifier or an oscillator, as outlined below. The modulator is then the tube and accompanying circuit which causes the bias to vary, and the tube in which the modulation products are produced is called a modulated amplifier.

Linear Modulation. It is possible to consider amplitude modulation from a different point of view from that used above--a point of view quite important if the carrier and modulating voltages are large, or if operation occurs near a sudden change in a tube characteristic, as at cutoff. If a carrier voltage, $E_0 \cos (\omega_0 t + \theta)$, were introduced into the plate circuit of a triode, the variation in potential caused by the signal voltage, $e_s = E_s \cos \omega_s t$, also introduced in the grid circuit could be considered the equivalent of a relatively slow change ($\omega_0 >> \omega_s$), in the grid bias. Assuming the bias to be near cutoff when e and e_s are absent, a plate current is produced which has a wave form indicated in Fig. 6-4. By use of a suitable output circuit, the carrier and the two side bands of amplitude modulation can be separated from the wave (a parallel resonant circuit tuned to carrier frequency may give large voltage at the carrier and side frequencies, small voltage at other frequencies), and hence a wave of form (6-4) is yielded.

The problem can be viewed somewhat more generally. It is almost always possible to set up a series approximating any curve as closely as desired. Equation (6-17) is such a series. If for a tube, whether diode, triode, tetrode or other, a dynamic characteristic can be obtained, that is a characteristic for actual a-c operating conditions making full allowances for changes in potentials of various electrodes when currents are varying, then a series can be set up to approximate this characteristic as closely as needed. Thus a curve of output current i versus input voltage e, on whatever electrodes these may appear, permits writing an equation

$$i = b_1 e + b_2 e^2 + \dots\dots \qquad (6\text{-}21)$$

where the b's are constants for a given tube under fixed operating conditions.

However, if e is not small, or if the series is not a simple Taylor series (when the tube is driven beyond cutoff the series will not be the Taylor series which is usually implied by 6-17), the use of the series may become quite involved. To simplify the problem, consider the case in which e is the sum of two sine-wave voltages

$$e_o = E_o \cos(\omega_o t + \theta)$$

$$e_s = E_s \cos \omega t$$

No matter how extensively the series (6-21) might have to be taken for practical purposes, it will be possible to break the result into $E_o \cos(\omega_o t + \theta')$ multiplied by a factor G involving e_s, the circuit parameters, etc., plus $\cos(2\omega_o t + \theta'')$

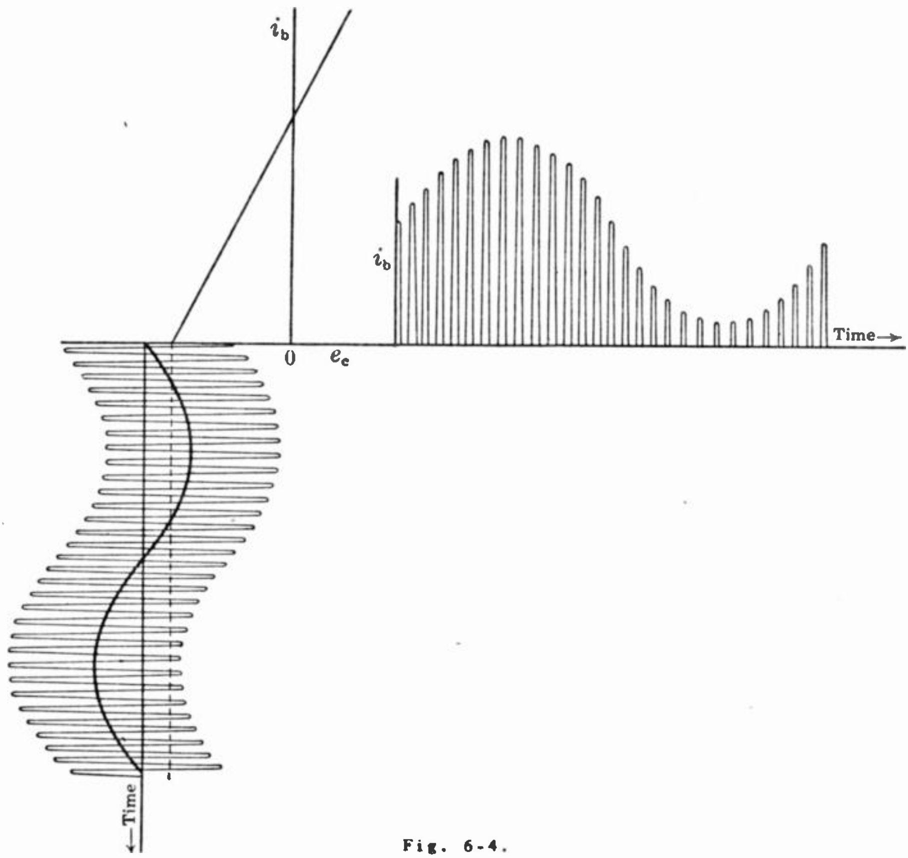


Fig. 6-4.

multiplied by some other factor, and so on. Hence

$$i = GE_o \cos(\omega_o t + \theta') \tag{6-22}$$

is one part of the response. Now if G varies linearly with e_s, so that

$$G = G_o(1 + kE_s \cos \omega_s t)$$

then this part of the response becomes

$$i = G_0E_0\left(1 + k\frac{E_s}{E_0}\cos\omega_s t\right)\cos(\omega_0 t + \theta')$$

$$= G_0E_0(1 + m_a\cos\omega_s t)\cos(\omega_0 t + \theta') \qquad (6\text{-}23)$$

which is an amplitude-modulated wave (see eq. 6-3).

In the case illustrated in Fig. 6-4, eq. (6-23) still holds, but G_0 must be interpreted as a constant at each instant for which $e_0 + e_s >$ cutoff, and zero at other times, thus clipping off the bottoms of the output wave.

From the above discussion the source of the name "linear modulation" should be clear, and we can see the desirability of having a linear G vs E_0 characteristic to reduce the output to (6-22), and thus eliminate distortion. But distortion may be tolerated if it can be eliminated, as is possible in the case of Fig. 6-4, and is accepted in such cases for the sake of greater efficiency.

The efficiency of the modulation can be increased when using a Class B or Class C driver, and the appropriate components in the plate current can still be separated out for use. Figures 9-3 and 9-4 (Ch. 9) show circuits for plate-circuit modulation and grid-circuit modulation. In Fig. 9-4 it can be seen that the portion of the circuit marked "modulator" is the part which varies the grid potential, in accordance with the input signal (indicated by "From Speech Amplifier"), of the tube in the part of the circuit labeled "Modulated Amplifier." The carrier (indicated by "Class C Driver") is introduced into the grid circuit of the "Modulated Amplifier" in the plate circuit of which is a tuned parallel circuit from which the voltage output is taken.

In Fig. 9-3 is shown a circuit for plate modulation (variation of plate potential by signal voltage instead of grid bias as above) of an amplifier. The signal (entering at terminals marked "From Speech Amplifier") feeds into a Class B push-pull circuit the voltage output of which is introduced into the plate circuit of the modulating amplifier. The plate potential of the latter is thus varied in accordance with the signal voltage, while the grid potential is varied in accordance with the carrier voltage input. A parallel tuned circuit separates the desired modulation voltages from those of other frequencies, and thus supplies the output.

6-6 Types of Frequency Modulation.- There are two important methods of obtaining frequency modulation, by reactance variation and by the Armstrong method.

Reactance-variation Method of Frequency Modulation. The reactance-variation method of frequency modulation consists essentially of varying the frequency of the output of an oscillator by introducing the equivalent of a varying capacitance or inductance into the oscillator circuit. It was shown in Ch. 4 that numerous common oscillators produce oscillations of frequency given approximately by

$$\omega = \frac{1}{\sqrt{LC}} \tag{6-24}$$

If now L remains constant and C varies so that

$$C = C_o(1 + k \cos \omega_s t) \tag{6-25}$$

where $0<k<1$ and $k \cos \omega_s t$ is directly proportional to a signal $e_s = E_s \cos \omega_s t$, then

$$\omega = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{LC_o(1 + k \cos \omega_s t)}} \cong \frac{1}{\sqrt{LC_o}}\left(1 - \frac{1}{2}k \cos \omega_s t\right) \tag{6-26}$$

the last approximation holding only when k is quite small. Since $1/\sqrt{LC_o}$ can be considered the carrier frequency (it is the value of ω when $k = 0$), the instantaneous frequency given by (6-26) is directly proportional to the signal e_s, hence frequency modulation is attained.

The above argument will serve as an introduction, but cannot be considered other than an extremely gross approximation. Equation (6-24) is derived on the assumption that L and C are constants, and does not hold when a circuit parameter such as C varies with time. It is possible, however, to set the theory of the reactance-variation method on a better foundation. The simplest oscillator is an L-C circuit (Fig. 6-5) containing either no resistance, or a device which counterbalances the effect of resistance by supplying energy to the circuit to compensate for both resistance losses and energy supplied to the load, if any. Many commor oscillators reduce to the equivalent of this.

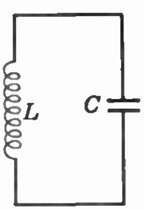


FIG. 6-5. Simple *L-C* circuit; when effective resistance of circuit is zero, represents simplest oscillator.

Let $i = dq/dt$ be the current in the circuit of Fig. 6-6. The total capacitance C in the circuit is $C_o(1 + k \cos \omega_s t)$, and

$$L\frac{di}{dt} + \frac{1}{C}\int i\,dt = L\frac{d^2q}{dt^2} + \frac{q}{C} = 0 \tag{6-27}$$

whence

$$\frac{d^2q}{dt^2} + \frac{q}{LC_o(1 + k \cos \omega_s t)} = 0 \tag{6-28}$$

Now $1/LC_o = \omega_o^2$ where ω_o may be considered the carrier frequency, since it is the frequency of oscillation when $k = 0$; and when $k << 1$, (6-28) may be written

$$\frac{d^2q}{dt^2} + \omega_o^2(1 - k \cos \omega_s t)q = 0 \tag{6-29}$$

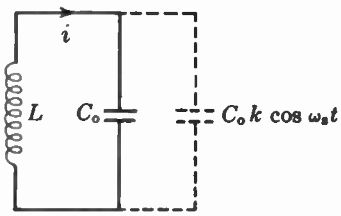


FIG. 6-6. *L-C* circuit with periodically varying capacitance. Variable portion indicated (dashed) as separate capacitor.

Substitution will show that so long as ω_0/ω_s is large (>> 1) and k is small (<< 1), an approximate solution is

$$q = \sin \int \omega dt$$

and

$$i = \omega \cos \int \omega dt$$

where

$$\omega = \omega_0 \sqrt{1 - k \cos \omega_s t}$$
$$= \omega_0 \left(1 - \frac{1}{2} k \cos \omega_s t\right) \quad (6\text{-}30)$$

Thus i has both amplitude and frequency modulation, and the "instantaneous frequency" is ω. Figures 6-7 and 6-8 show the actual q and i which are solutions of (6-28)

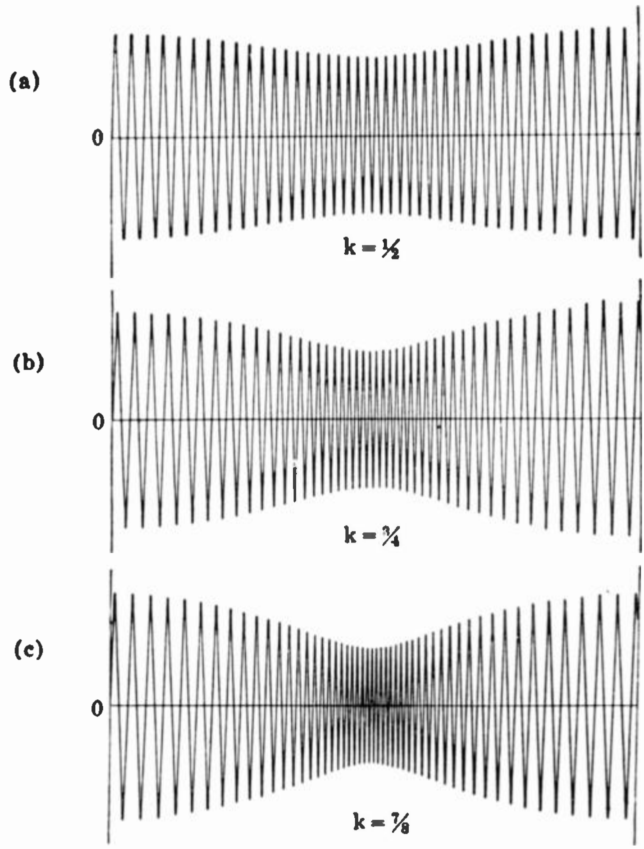


FIG. 6-7. Variation with *k* of charge *q* in eq. (6-28); $\omega_o/\omega_s = 40$.

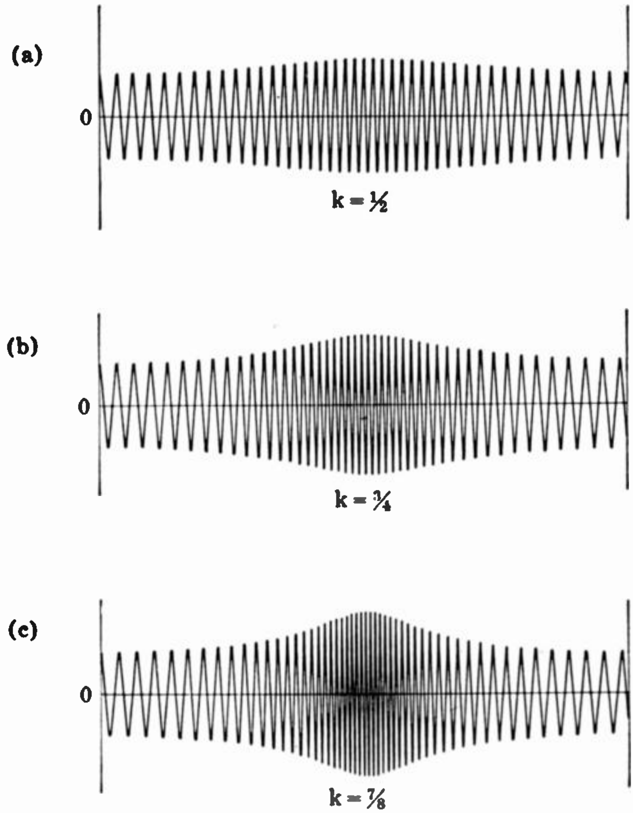


FIG. 6-8. Current $i = dq/dt$ in eq. (6-28); $\omega_0/\omega_s = 40$; k as indicated

when k is not small. The figures serve to illustrate a frequency-modulated wave with some amplitude modulation, and the result which may be expected when such a wave is supplied to a limiter is illustrated in Fig. 6-9. A limiter is a device to limit

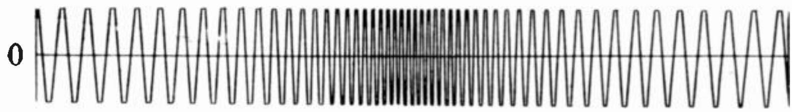


FIG. 6-9. Current $i = dq/dt$ in eq. (6-28); $\omega_0/\omega_s = 40$; k as indicated.

amplitude, and one of the advantages of frequency modulation is that amplitude modulation can be clipped off the wave without substantial harm to the frequency modulation.

The production of the equivalent of a varying capacitance or inductance may be obtained by a tube circuit of which the typical elements are shown in Fig. 6-10a, in which L and C_o are the fixed parameters of an oscillator circuit, $\underline{Z}_1$ and $\underline{Z}_2$ are two impedances, in series and with the same current $\underline{I}_1$ in each so long as no appreciable grid current flows. Assuming the tube to be an amplifier, the equivalent plate circuit theorem yields Fig. 6-10b as the equivalent circuit. The admittance $\underline{Y}_{AB}$ look-

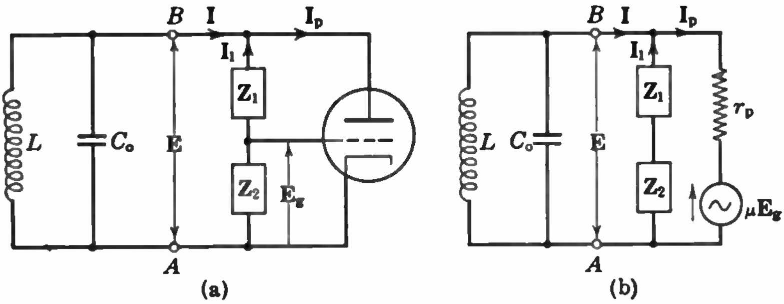


FIG. 6-10. Reactance tube circuit and equivalent circuit.

ing into the terminals AB to the right is

$$\underline{Y}_{AB} = \frac{\underline{I}}{\underline{E}} \doteq \frac{g_m \underline{E}_g}{\underline{E}} \qquad (6\text{-}31)$$

where the last expression is an approximation.

For $\underline{Z}_1 = 1/j\omega C_1$ and $\underline{Z}_2 = R_2$

$$\underline{Y}_{AB} \doteq g_m \frac{\underline{E}_g}{\underline{E}} = g_m \frac{\underline{Z}_2}{\underline{Z}_1 + \underline{Z}_2} \doteq g_m \frac{\underline{Z}_2}{\underline{Z}_1} = j g_m \omega C_1 R_2 \qquad (6\text{-}32)$$

Thus the impedance looking into AB approximates that of a capacitor of capacitance $g_m C_1 R_2$. If, now, by inserting the modulating signal in the biasing circuit of the grid the mutual conductance g_m can be varied in accordance with the modulating signal, then the capacitance looking into AB will be varying according to the modulating signal, and this will frequency-modulate the oscillator. A different combination from that used above is $\underline{Z}_1 = R_1$ and $\underline{Z}_2 = 1/j\omega C_2$, which yields the equivalent of a varying inductance. This circuit is shown in Fig. 8-5 of Chapter 8 where, however, the reactance tube is being used in an automatic frequency control[12]--the potential from the discriminator in the

12. Travis, Charles, "Automatic Frequency Control," Proc. I.R.E., Vol. 23, No. 10, Oct. 1935; D.E. Foster and S.W. Seeley, "Automatic Tuning, Simplified Circuits, and Design Practice," Proc. I.R.E., 25, No. 3, March (1937).

figure changes the grid bias, hence the g_m of the tube, hence the frequency of the oscillator.

It will often be convenient to apply the oscillator voltage to one grid and the signal voltage to another; in Fig. 6-11

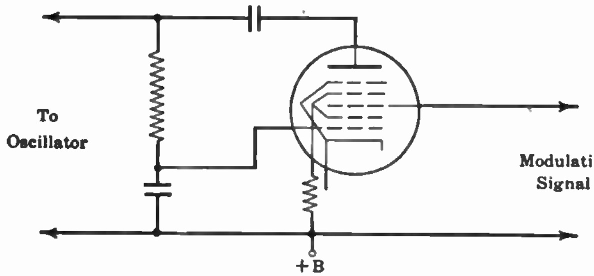


FIG. 6-11. Pentagrid tube used as reactance tube.

the oscillator input and the modulating signal circuits are shielded from one another.

Armstrong Method of Frequency Modulation.[13] The Armstrong method achieves frequency modulation by way of amplitude modulation and phase modulation. By (6-3)

$$e_a = E_a(1 + m_a \cos \omega_s t) \cos (\omega_o t + \theta)$$

$$= E_a \cos (\omega_o t + \theta) + m_a E_a \cos \omega_s t \cos (\omega_o t + \theta) \qquad (6\text{-}33)$$

is an amplitude-modulated wave. Now assume that a 90° phase shift can be introduced in the $\cos(\omega_o t + \theta)$ in the second term of (6-33). Then

$$e_{ap} = E_a \cos (\omega_o t + \theta) + m_a E_a \cos \omega_s t \sin (\omega_o t + \theta)$$

$$= E_a\sqrt{1 + m_a^2 \cos^2 \omega_s t} \cos [\omega_o t + \theta - \tan^{-1} (m_a \cos \omega_s t)] \qquad (6\text{-}34)$$

is the result, and is amplitude modulated[14] since the amplitude $E_a\sqrt{1 + m_a^2 \cos^2 \omega_s t}$ contains the signal, and is also phase modulated, since the varying phase angle

$$\tan^{-1}(m_a \cos \omega_s t) \qquad (6\text{-}35)$$

13. Armstrong, Edwin H., "A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation," Proc. I.R.E., 24, No. 5, May (1936). Joffe, D.L., "Armstrong's Frequency Modulator," Proc. I.R.E., 26, No. 4, April (1938).

14. But the amplitude modulation is of a different type from (6-3) because of the square root and the square.

also contains the signal. For small values of m_a, the arc tangent may be replaced by its argument $m_a \cos \omega_s t$, and by suitable limiter circuits the varying amplitude, which is almost E_a times unity when m_a is small, may be made constant. The e_{ap} wave of (6-34) then becomes approximately the phase-modulated wave

$$e_p = E_a \cos (\omega_o t - m_a \cos \omega_s t + \theta). \qquad (6\text{-}36)$$

It will be recalled from (6-2) that $m_a = kE_s/E_o$ where E_s and E_o are maximum values of the sine-wave signal and carrier respectively and k is a constant, and it will further be recalled from (6-10) that a frequency-modulated wave requires m_f/ω_s in place of m_a in (6-36). This can be attained by supplying to the modulator, not the signal voltage of magnitude E_s, but a voltage proportional to E_s/ω_s. The latter is obtained by sending a sine-wave current i_s, directly proportional to the signal voltage e_s, through an R-C series circuit. The voltage across the capacitor C is directly proportional to $I_s/\omega_s C$, and hence to E_s/ω_s. Using this voltage instead of e_s to modulate the carrier, frequency modulation results. But since the requirement that m_a in (6-35) be small to obtain (6-36) limits the frequency variation about ω_o to an undesirably small amount, it is customary to carry out the process outlined above at relatively low frequencies, and then use frequency multipliers to increase the frequency. Since such multipliers multiply the entire phase of a sine wave, the frequency deviation is increased although its ratio to the carrier at any stage is the same as at any other.

The essentials of the Armstrong system of frequency modulation are outlined in Fig. 9-5 of Chapter 9. The signal $e_s = E_s \cos \omega_s t$ to be transmitted is fed to the amplifier marked "Audio Amplifier," in the plate circuit of which is a series R-C circuit so proportioned that the magnitude of the sine-wave voltage across C is directly proportional to E_s/ω_s. This voltage is fed to the balanced modulator. A crystal-controlled oscillator supplies a voltage of frequency of the order of magnitude of 200 kc to the balanced modulator through the buffer amplifier, which serves to isolate the oscillator from the rest of the circuit and thus avoid changes in oscillator frequency with changes in the other parts of the circuit. The balanced modulator, one form of which is illustrated in Fig. 6-12, balances out currents of carrier frequency in the plate circuit but passes the side frequencies, as explained below. By making the plate circuit non-reactive and using a transformer to supply the output, a 90° phase shift is introduced. The phase-shifted side bands are then added to the oscillator output in a combining amplifier. Since the sum is of the form of (6-34), a frequency-modulated output leaves the combining amplifier. The need for frequency multipliers, shown following the combining amplifier, has already been explained. The converter following will be discussed in Ch. 7.

The balanced modulator shown in Fig. 6-12 is of interest. The oscillator supplies voltage of the same phase to each grid.

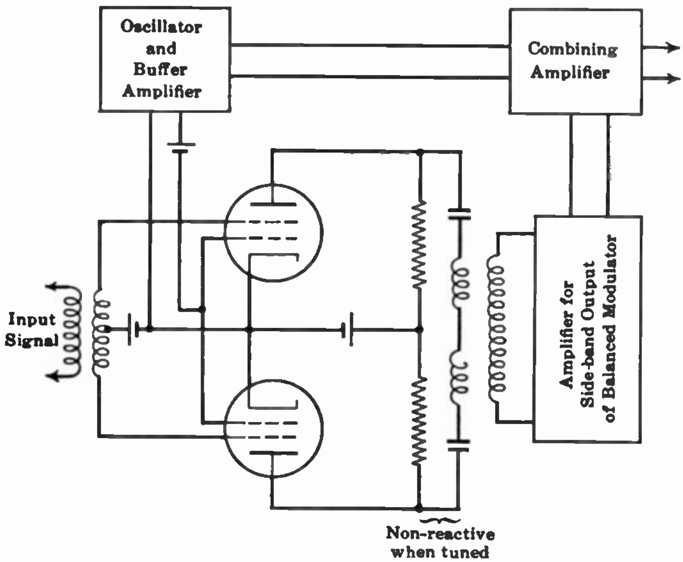


FIG. 6-12. Balanced modulator and 90° phase shift circuit for Armstrong system.

The input signal is supplied to the second grids of the tetrodes in push-pull fashion. Applying small-signal (square-law) analysis as given under amplitude modulation (Sec. 6-5), it can be seen that first order terms of carrier frequency drop out, whereas second order remain. Thus if the oscillator or carrier voltage is $e_0 = E_0 \cos \omega_0 t$ and the signal voltage $e_s = E_s \cos \omega_s t$, the output plate current, for non-reactive load, will be of the form $a_1 e + a_2 e^2 + \ldots..$ where $e = e_0 + ke_s$ (k = constant) for one tube and $e = e_0 - ke_s$ for the other. Hence the net current in the plate circuit will be

$$[a_1 e_0 + a_2(e_0 + ke_s)^2 + \ldots] - [a_1 e_0 + a_2(e_0 - ke_s)^2 + \ldots]$$

or $4a_2 ke_0 e_s$, which contains just the two side-frequency terms, without any term of oscillator frequency.

6-7. Modulation at High Frequencies.- Modulation at the higher frequencies is not greatly different from that at lower frequencies (less than 200 Mc). Transit time becomes important, but its adverse effects may in some cases be overcome by compound modulation, that is, by applying the modulating voltage at several places in the circuit so that undesired effects are cancelled. For example, Lindenblad[15] has reported obtaining amplitude modulation "substantially free from frequency variations" (which in A-M constitutes a distortion) by modulating in phase, and in experimentally determined relative magnitudes, both the grid and the plate of a positive-grid oscillator (Ch. 10). Increasing the positive grid potential increases the electron velocity, and a simultaneous decrease of plate potential (already negative) increases the length of path, leaving the frequency of oscillation of the electron about the grid constant. Thus the amplitude of the output may be varied while the frequency remains fixed, yielding an amplitude-modulated wave. Without the compound modulation, F-M would be introduced in addition to A-M.

It is possible to modulate at relatively low frequencies and by the use of frequency multipliers to step up the frequency. This scheme has already been illustrated in the Armstrong system of F-M, where however, it is used to keep (6-35) small rather than because of the size of the channel or transmitting frequency range.

High-frequency devices such as velocity-modulated oscillators, discussed in Ch. 10, may be modulated directly.[16] Equation (10-23) of Ch. 10 shows that the frequency of a klystron varies with applied voltage in relatively simple fashion, so that F-M is not difficult to obtain. But the very fact that the frequency of the klystron responds so readily to applied voltage makes A-M difficult to obtain, since F-M is quite likely to be present and thus distort the A-M.

---

15. Lindenblad, N.E., "Development of Transmitters for Frequencies above 300 Megacycles," Proc. I.R.E., 23, No. 9, Sept. (1935). Compound modulation may be useful at lower frequencies--see C-Y Meng, "Linear Plate Modulation of Triode Radio-Frequency Amplifiers," Proc. I.R.E., 28, No. 12, Dec. (1940).

16. Note the double use of the word modulated in this sentence. "Velocity-modulated" refers to change in the velocity of electrons in a previously uniformly moving stream, and is used in an oscillator, for example, in which no modulation of the type discussed in this chapter is desired.

Chapter 7

DEMODULATION (DETECTION)

Demodulation, or detection as it is usually called, is the recovery from the modulated wave of the original signal that modulated the carrier at the transmitter. In most cases a replica of the original modulating signal is desired, and the deviation from the original form in the demodulated wave constitutes distortion. In other cases an indication, but not necessarily a replica, of the original signal is wanted; and in still other cases detection consists of eliminating the original carrier from the modulated signal that arrives at the receiver, and substituting therefor a new carrier suited to certain circuits of the receiver. It is customary to consider the last case as detection, although it may also be looked on as modulation.

The following types of detection are prominent:

Small-signal or square-law detection
Large-signal or linear detection[1]
(a) "average" detection
(b) envelope detection
Automatic volume control
Mixers and converters
Detection of frequency-modulated waves.

These items will be considered seriatim.

7-1. Small-Signal, or Square-Law, Detection.- By small-signal or square-law detection is meant detection characterized by (1) small input signal, so that in any cycle of the modulated input, currents and voltages vary over relatively small ranges of the characteristics, and (2) a curved characteristic relating the output current i in any electrode to the input modulated voltage e introduced into any electrode circuit. Under these circumstances

$$i = a_1 e + a_2 e^2 + \ldots.. \qquad (7\text{-}1)$$

where the a's are constants, depending on the circuit parameters and the parameters of the tube at the operating point.

Now let e be an amplitude-modulated wave

$$e = E_a \,(1 + m_a \cos \omega_s t) \cos (\omega_0 t + \theta)$$

1. Not all large-signal detection is necessarily linear detection, particularly in measurements. See footnote 10.

$$= E_a \cos(\omega_o t + \theta) + \frac{1}{2} m_a E_a \cos[(\omega_s + \omega_s)t + \theta]$$
$$+ \frac{1}{2} m_a E_a \cos[(\omega_o - \omega_s)t + \theta] \qquad (7\text{-}2)$$

Then

$$e^2 = 2\, m_a E_a^2 \cos \omega_s t + \frac{1}{2} m_a^2 E_a^2 \cos 2\, \omega_s t \qquad (7\text{-}3)$$

+ terms of twice carrier and twice side-band frequencies

+ d-c terms,

and i (7-1) will have terms of angular frequency ω_s, $2\omega_s$, as well as others of frequencies lying in the carrier range and higher. Thus by appropriate filtering in the circuit in which the current, i, flows, the original signal (ω_s) can be obtained, but distortion is also present ($2\omega_s$) and other terms of the series of (7-1) may introduce further distortion terms. The other terms in the series depend on still higher powers of e, and their effect may be minimized by keeping e small. Hence the name small-signal detection. Since the desired result proceeds from the e^2 term in (7-1), the alternative name square-law detection is also appropriate. The reader will find it instructive to determine the response in the frequency region of the original signal when e is an amplitude-modulated wave containing two (or more) original signal frequencies.[2]

The coefficient a_2 in (7-1) is important, in that the desired result is directly proportional to it. By properly controlling the circuits of the detector and adjacent stages, it may be made relatively large, thus producing a gain over the result obtained for smaller values of a_2. The evaluation for numerous special cases of a_2, or its equivalent, will be found in the literature.

In the above discussion, i and e have been specified as the current in a particular electrode and the voltage introduced in the same or some other electrode. In a practical detector, a wide variety of different connections are in use. For example, a connection that makes use of the I_b - E_b characteristic would mean that e is introduced in, and i flows in, the plate circuit. Some of these possible arrangements for small-signal detection are given below:

1. Diode: I_b - E_b characteristic; total plate voltage e_b always greater than zero.

2. Since the desired term of (7-3) can be obtained from (7-2) only when the carrier wave as well as the side-frequency terms are present, transmission of only side-frequency terms requires that the carrier be re-supplied at the receiver. However, carrier and one side-frequency term are sufficient to produce an ω_s term in e^2, so that it is not necessary to transmit both side-frequency terms.

2. Triode, or multielectrode tube in which all electrodes except one grid and the plate are unaffected by the input e:

(a) $I_b - E_b$ characteristic (grid always negative).

(b) $I_b - E_c$ characteristic (grid always negative; often called plate detection although this is a misnomer[3]).

(c) $I_c - E_c$ characteristic (grid always positive; dynamic $I_b - E_c$ characteristic straight; this is commonly called grid detection[4]).

The case of 2(c), grid detection, can be used as an example. The coefficient a_2 will be different for each term of different frequency in the sum of sine waves to which e^2 can be reduced. By carrying out the analysis of the circuit, it can be shown that the magnitude of a_2 for the term cos $\omega_s t$ is[5] given by

$$\frac{\mu}{D} \frac{Z_s}{[r_g + Z_c]^2 \; [r_g + Z_s] \; [r_p + Z_{Ls}]} \tag{7-4}$$

where D is a quantity known as the detector voltage constant which depends on the curvature of the $I_c - E_c$ characteristic (and is infinite if this characteristic is a straight line), Z is the magnitude of the impedance in the grid circuit, Z_L is the magnitude of the impedance in the plate circuit, $r_g = \partial e_c/\partial i_c$ (similar to $r_p = \partial e_b/\partial i_b$), and subscripts s and c indicate that the impedance is to be evaluated either at the angular frequency of the original modulating signal ω_s, or at the angular frequency ω_o of the carrier.

To make (7-4) large, μ and Z_s should be large, Z_c, Z_{Ls}, and D should be small. So far as the impedances are concerned, this result is achieved in the circuit of Fig. 7-1. The $R_k - C_k$ combination in the external grid circuit makes Z high for ω_s but low for ω_o; and the capacitor C_L helps to make Z_L high at ω_s, and by-passes the currents of carrier and higher frequencies, but has high impedance to lower frequency components, so that the desired current components of the frequency of the original modulating signal pass through the coil.

It is not possible here to go into other circuits for small-signal (square-law) detectors, but it will be appreciated

3. Also called "anode-bend" detection, which may be bottom-bend (triodes) or bottom-bend or top-bend (pentodes); the "bend" refers to the curvature of the characteristic which may be near top or near bottom.

4. The detected current actually appears in the grid circuit, but is usually extracted in the plate circuit. This leads to an analogy in which the cathode and grid are considered a diode, and the plate is the equivalent of a triode to the grid of which the signal from the equivalent diode is fed.

5. Cited from McIlwain and Brainerd, "High Frequency Alternating Currents," p. 229; texts such as those of Reich, Terman, Everitt, Eastman, Glasgow, etc., contain discussions.

that the basis for the successful operation of such detectors comprises the following points: (1) a curved characteristic having an e^2 term, (2) a small signal to keep down distortion, and (3) an adequate circuit analysis to permit design for maximum gain and minimum distortion.

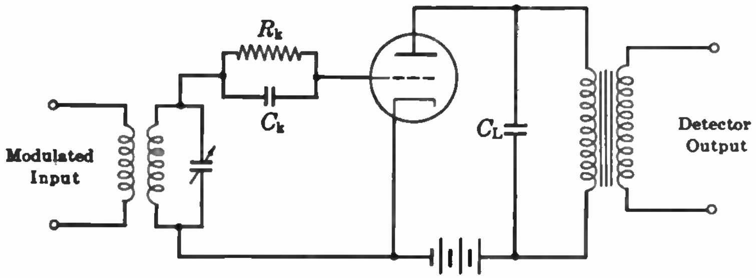


FIG. 7-1. Circuit for Grid Detection.

7-2. Large Signal, or Linear, Detection.- As implied by the name, large-signal detection occurs when the modulated input to the detector is large. This type is the converse of the small-signal detector of the previous section, and the name arises from the fact that the magnitude of the modulated wave[6] input must be large if distortion is to be kept small. On first consideration, this statement may seem to contradict those of the previous section, but actually it does not because of the different conditions under which large-signal detection takes place. The large magnitude of the signal permits such a great portion of a characteristic to be used that the response is substantially linear, even though part of the characteristic is curved. If the tube is biased at cutoff, or otherwise made nonconducting for at least half of each cycle of the input-modulated wave, the original modulated signal is detected as will be discussed below.

Average Detection. Although not so common or so efficient as envelope detection, average detection will be considered first because of its simplicity. If a tube--frequently a diode--is biased at cutoff, and a pure resistance load or its equivalent connected in the plate circuit, then the tube will act as a

6. Ten volts (peak value) is sometimes considered a minimum in commercial broadcast receiver design; it is convenient to consider that two volts is the boundary between a small signal and a large signal.

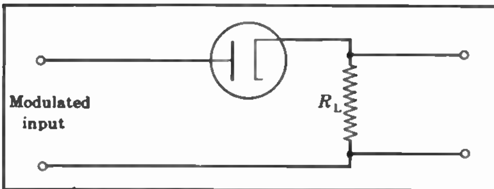
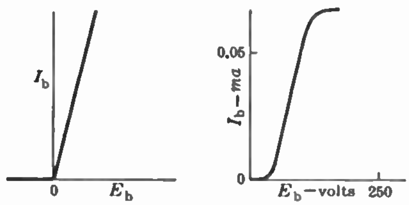
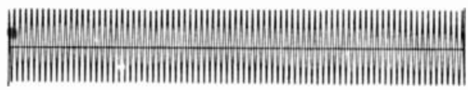
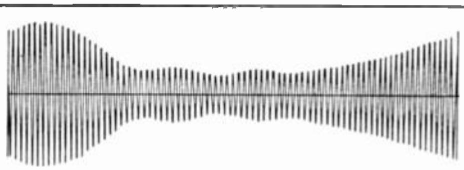
	(a) Diode detector "average" type.
	(b) Left – diode characteristic considered ideal for many purposes. Right – an actual static diode plate characteristic.
	(c) Carrier alone.
	(d) Form of current in diode circuit when carrier alone is applied (ideal case; left characteristic of (b) above).
	(e) Form of average component of current in (d).
	(f) Amplitude-modulated wave.
	(g) Form of current in diode circuit when amplitude-modulated wave is applied.
	(h) Form of average component of current in (g).

FIG. 7-2. "Average" large-signal detection.

rectifier[7] and only the half of a cycle "of one polarity" will cause a current to flow. Hence, if an unmodulated carrier is impressed (Fig. 7-2) on a diode, a current will pass through R_L that has an average[8] value proportional to E_m/π where E_m is the maximum value of each half-cycle of the carrier wave.[9] Thus the current in the plate circuit is composed of a direct current (the average value) plus components of carrier frequency.

If now an amplitude-modulated wave is impressed on the tube (Fig. 7-2f) and if the carrier is of a much higher frequency than that of the original signal, the average value of the current over any short period of time will be proportional to E_m/π where E_m is the peak value of the half-cycles in the shorter period. E_m now varies as $E_a(1 + m_a \cos \omega_s t)$, and hence there is a component of current that also varies in this manner; that is, there is a component reproducing the original modulating signal, and by appropriate circuits it may be separated from the other components.

The above reasoning is based on the assumptions that the diode or other device has a constant, finite resistance to voltages of "one polarity" and an infinite resistance to voltages of opposite "polarity." Thus the ideal characteristic from this point of view is that shown in the left of Fig. 7-2b. Diodes are often used because they closely approximate a linear characteristic when the impressed voltage has a sufficiently large magnitude to make the (inevitable) curvature near cutoff small in comparison with the substantially straight part of the characteristic over which the voltage varies.[10]

---

7. The distinction between diode rectifiers and diode detectors is not always sharp. In this text "diode detector" and "diode rectifier" are terms based on use of the diode. Some persons refer to diode detection as rectification. See for example H.A. Wheeler, Design Formulas for Diode Detectors, Proc. I.R.E., 26, No. 6, June (1938).

8. The result is worked out in Ch. 1, section on Fourier series.

9. On a voltage basis, the maximum efficiency (ratio of useful voltage across load to input peak voltage) is $1/\pi$ for a single diode under ideal conditions for average detection.

10. It is well to bear in mind that although most detection problems have for an objective the obtaining of a replica of the original modulating signal, or, if there has been distortion preceding the detector, at least a replica of the signal contained in the modulated wave, nevertheless numerous problems arise, particularly in measurements, where a replica is not essential so long as a calibration curve or other equivalent characteristic of the detector is known.

It is also desirable to keep before one the fact that a "crystal" with which a wire or piece of metal is in contact may have a characteristic quite similar to that of a diode, and hence may under some circumstances be used in place of a diode. At ultra-high frequencies this is particularly true, and crystals detectors have found a new and extensive field of application in this region.

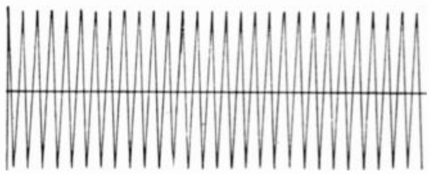
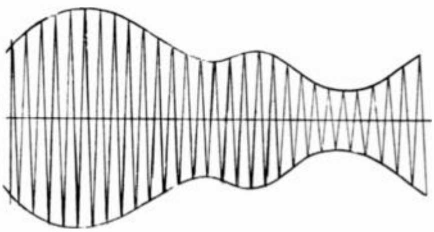
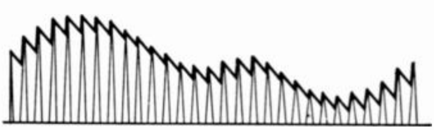
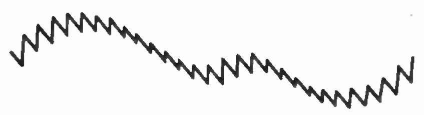
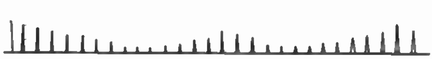
	(a) Carrier. (Original carrier or i-f carrier of last i-f stage).
	(b) Amplitude-modulated wave with envelope outlined; see next entry (c).
	(c) Envelope of the amplitude modulated wave; this is the desired signal, modified by distortion and noise introduced in transmitter, in transmission, and in receiver up to detector.
	(d) Illustration of the charging and discharging of the capacitor C of Fig. 7-4 (a).
	(e) Instantaneous variation of voltage across the capacitor C, showing desired signal (see (c) above) plus carrier-frequency pulses which must be filtered out.
	(f) Portions of carrier wave during which current flows; illustrates high efficiency obtainable by this method of detection (on power basis, small currents for short times; on voltage basis maximum detected signal voltage can be made to approximate peak input voltage).
	(g) Illustration of introduction of distortion by incorrect R-C circuit design. The light curve shows the signal (as in (c) above); the voltage across the capacitor does not follow the envelope (desired signal) at D because it does not fall rapidly enough to respond to the peaks of the input wave (b).

FIG. 7-3. Envelope Detection.

Envelope Detection. In envelope detection, more customary than average detection because of the greater efficiency obtainable, a diode or other tube is unbiased or is biased at cutoff. The output however is not an average of half-cycle waves, but a voltage which, to a very close approximation, varies as the envelope of the amplitude-modulated wave because of the presence of an R-C circuit in the plate circuit (Fig. 7-4a).

The operation is of this nature: each half-cycle "of positive polarity" of the amplitude-modulated wave (Fig. 7-3b) applied in the tube circuit will cause current to flow provided the plate remains positive. But as soon as a condition closely corresponding in detection to the ordinary steady state of linear circuits has been reached, the capacitor C remains charged (hence has a voltage across it) at the end of a cycle. Thus the input voltage must first balance this capacitor voltage before the plate again becomes positive with respect to the cathode. Consequently, with proper values of R and C, plate current flows only during the positive peaks of the applied voltage. When no current flows the capacitor discharges, and when current flows the capacitor is charged. In order to produce a voltage closely resembling the envelope of the applied amplitude-modulated signal, the drop during any cycle of the voltage across the condenser must be sufficient so that it may be raised again on the next following cycle. When this condition does not obtain, distortion is introduced. Figure 7-3 illustrates the steps in envelope detection and Fig. 7-4 shows several circuits. It is essential, if the graphical analysis of Fig. 7-3 is to hold without modification, that the effective R-C circuit of the diode not be shunted to any great extent. This restriction, and the need for large input voltage (the reader can show from Fig. 7-3 that this is essential), are the primary design limitations on diode detectors.

Figure 7-4 snows a number of diode detector circuits, which serve also to illustrate the use of multi-purpose tubes.

Rectification and Transrectification Characteristics. A large amount of information concerning the operation of a detector can be gained from curves known as the rectification and transrectification characteristics. According to I.R.E. standards,[11] a rectification characteristic is one in which "the average currents in an electrode circuit, as read by a d-c instrument, are plotted as ordinates against values of the direct voltage ... on the electrode as abscissas, for various values of E as a parameter; i.e., E is held constant for each graph." In this statement, E refers to either the rms or the peak value of a sine wave alternating voltage input, which may be pictured as

11. Standards on Electronics, 1938, p. 47, I.R.E.

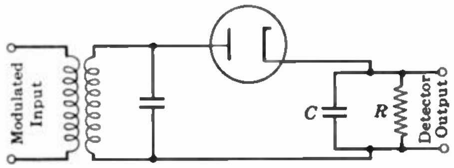
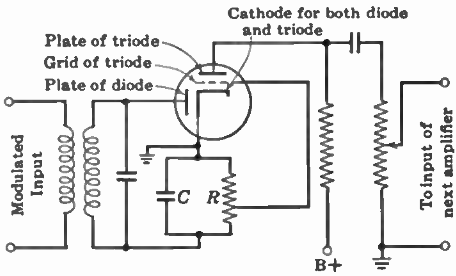
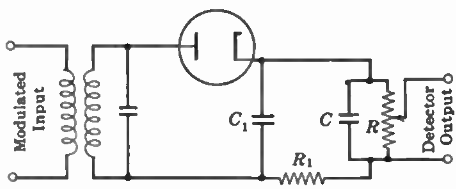
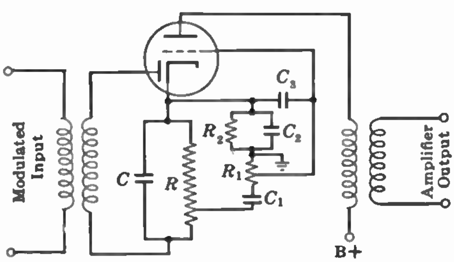
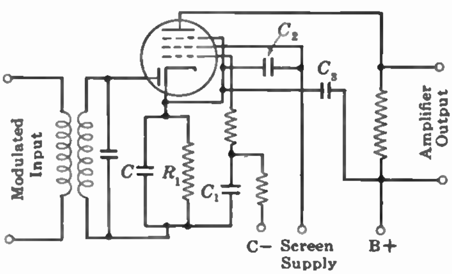
	(a) Simplest diode detector circuit.
	(b) Diode detector circuit of (a) and amplifier stage, using one tube (diode-triode, or duplex-diode-triode with plates of diodes connected together; one filament only); sometimes called diode-biased circuit because of approximately constant bias on grid of triode part when carrier only is input. Unless carrier is same value for different inputs, bias will vary.
	(c) Circuit of (a) with R_1 and C_1 added to reduce r-f in output; R_1 must be $\ll R$ (R_1 of the order of $0.1R$); C_1 can be of same order of magnitude as C.
	(d) Diode detector circuit of (a) and amplifier stage using one tube, as in (b); bias on grid of triode obtained by varying potential of cathode by drop through R_2, hence sometimes called cathode-biased circuit. C_2 shunts a-f currents; C_1 keeps bias from grid of triode; C_3 by-passes r-f currents which might reach grid. The basic R-C circuit must not be shunted appreciably.
	(e) Diode detector circuit of (a) used with pentode amplifier (diode pentode or duplex-diode pentode tube); fixed biases. C_1 keeps C bias from cathode, C_2 and C_3 pass r-f or i-f components from screen and output circuits to cathode.

FIG. 7-4. Examples of Diode Detector Circuits.

the carrier although low-frequency a-c may be used when inter-electrode effects and the like are not important.

A transrectification characteristic is a curve showing the relationship between "the average current in the circuit of an electrode, the direct voltage on that electrode, and the amplitude (or rms value) of an alternating voltage impressed on another electrode."

A diode has only a rectification characteristic. An example of one is shown in Fig. 7-5. If there is a d-c load in

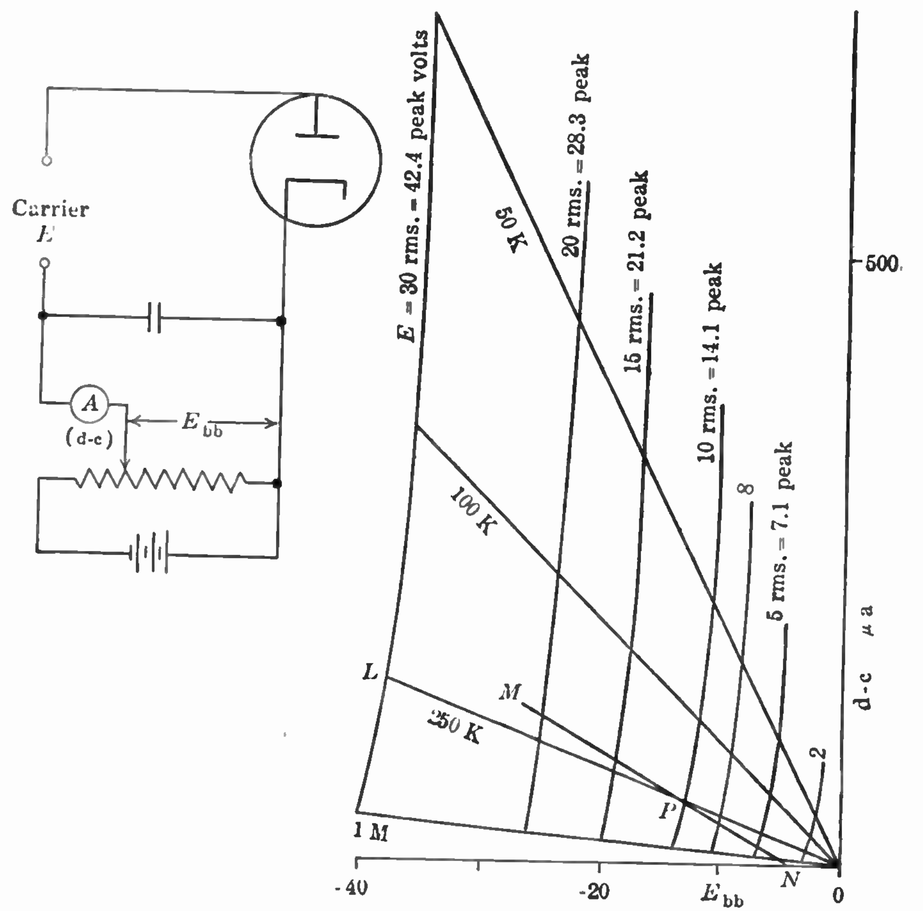


FIG. 7-5 Rectification characteristic of 6 H 6 diode, with load lines (K = 1000 ohms).
Each of the vertical curves has both the constant rms and the corresponding constant peak voltage of the carrier indicated

the circuit, i.e., a resistance R_L or its equivalent (Fig. 7-2a) the flow of d-c. through this resistance will cause a drop in the plate potential. Thus, if R_L = 0.1 megohm and the rectification diagram indicates a d-c. of 200 μa, a -20 volt potential would appear on the plate. For a fixed resistance R_L, a load line can be drawn, as in the case of the simpler tube characteristics discussed in Ch. 2. The line LO of Fig. 7-5 is a load line for R_L = 0.25 megohm; for any value of d-c plate current it shows the corresponding steady potential applied to the plate. (The line makes an angle $\cot^{-1} R_L$ with the negative abscissa axis.)

As an example, consider an impressed amplitude-modulated wave of 10 volts rms carrier 100 per cent modulated, with 0.25 megohm total resistance. For carrier alone, 52 μa d-c. would flow in the plate circuit and the operating point would be at P. For the modulated wave the operating point may be considered to move at the frequency of the original modulating signal along LO. The path would extend from 0 to the intersection of LO with the line for E = 20 rms volt. By plotting corresponding values of current against time, the slowly varying part of the output can be determined. Note that although LO is a straight line, the distance between the curves for various E's is not proportional to the E's, hence some distortion is introduced. Higher carrier voltage reduces the distortion.

In average detection with a pure resistance load, the detected output can be determined in advance. For envelope detection, a load line taking into account the a-c load (i.e., the resistance of the load) can be drawn. Without going into details concerning the procedure, it is sufficient to say that such loading may result in a load line such as MN (Fig. 7-5), from which it is seen that a sweep of 10 volts along MN from the operating point P carries the current to cut-off in the region NO, introducing severe distortion. Figure 7-6 shows the character of the variation of distortion with the per cent modulation for two different

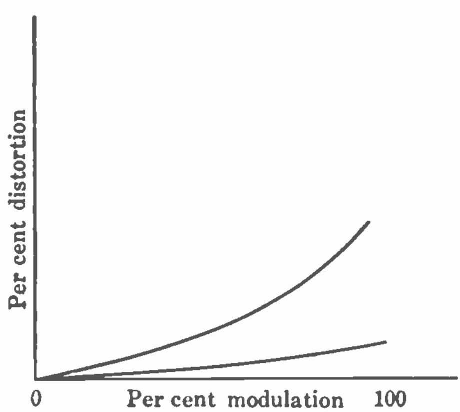


Fig. 7-6. Variation of distortion with per cent modulation; lower curve is for R-C circuit alone; upper curve for R-C circuit shunted by another circuit. (See Smith, Radiatron Designer' Handbook, p. 161.).

values of load in a diode detector.

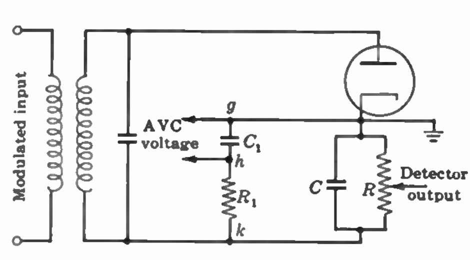
7-3. Automatic Volume Control (AVC).- The purpose of automatic volume[12] control (avc) is so to control a receiver that received modulated signals of far different carrier strengths will give outputs of about the same magnitude rather than of tremendously different magnitudes. AVC is closely allied with detection and for that reason is treated at this point.

There are two common types of avc, namely: simple, and delayed. Delayed avc does not mean a time delay, but a "delay" of the automatic gain control action until a certain minimum carrier voltage is obtained, thus permitting weak signals to be reproduced with the full gain of the receiver. An avc system ordinarily operates from a voltage proportional to the carrier and uses this voltage to bias one or more preceding stages. The bias changes to reduce the gain when the carrier strength increases, and vice versa.

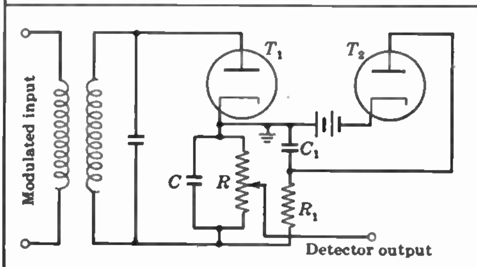
Simple AVC. A bias voltage for simple avc can be obtained from a diode detector, as shown in Fig. 7-7a. The R-C circuit of the detector retains the same form with, as without, the avc. The series R_1-C_1 circuit is shunted across it. Recalling that the positive direction of current flow is opposite to that of electron flow, point k will be at a lower potential than g (ground). The same voltage appears across R_1-C_1 as across R, and its form is approximately that of the envelope of the modulated input, but raised above the axis. In a properly designed detector, the actual voltage is at all times well above the axis. (See Fig. 7-3.) If C_1 is made such that the voltage across it cannot respond to changes in the envelope, but does respond substantially to the average of the voltage across R over any time long in comparison with the average period of the envelope curve, then the voltage across C_1 will be substantially proportional to the carrier of the modulated input, or to the equivalent in the case of disturbances of appreciable extent, and the point h will be at a potential negative with respect to the ground g. Thus the voltage across C_1 can be used to bias negatively any preceding stages, whether r-f, i-f (see Sec. 7-4), or other frequency. The magnitude of the bias will be closely proportional to the carrier strength. The choice of the magnitude of C_1 should be such that the time constant of R_1C_1 is so large that the charge on C_1 (hence also the voltage across it) cannot change greatly in any "cycle" of the detected voltage.

Delayed AVC. If it is desired to have the automatic volume control operate only after the carrier strength reaches

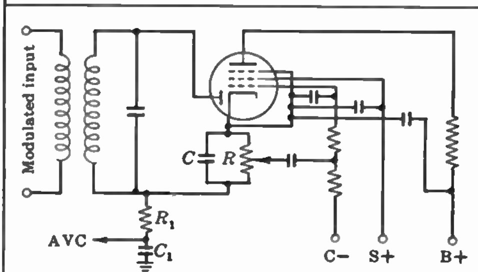
12. Not to be confused with a manual volume control, which may be obtained by making variable by hand or mechanically the connection from R marked "Detector output" in Fig. 7-7.



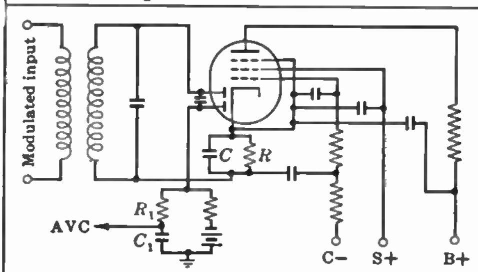
(a) Illustration of simple avc using diode detector.



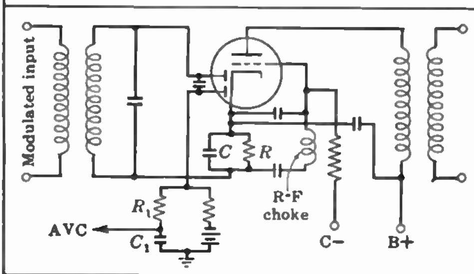
(b) Illustration of delayed avc (a twin diode with two plates and separate filaments in one envelope would usually be used).



(c) Diode pentode detector, simple avc, amplifier circuit (In case of duplex-diode pentode, the two diode plates are connected together.)
A diode triode circuit is substantially similar.



(d) Detector, delayed avc, amplifier circuit using duplex-diode pentode, (duplex-diode triode circuit is substantially similar—see next figure.)



(d) Detector, delayed avc, amplifier circuit using a duplex-diode triode; see (d) above.

FIG. 7-7. Some Examples of Automatic Volume Control.

a specified minimum (so that the reception of weak signals will not be affected by the avc system) a scheme such as that shown in Fig. 7-7b may be used. Diode T_1 and the R-C circuit comprise the detector; the series circuit R_1-C_1 supplies the avc voltage (across C_1) as in simple avc. However, a direct current from the battery flows in diode T_2 until the steady drop through R and R_1 (and through T_2, which is considered negligible) balances the voltage from the battery. At this point, a nearly steady voltage first appears across C_1 and may be used for the negative avc bias for preceding stages.

Figure 7-7 contains several avc circuits, which may be analyzed by the student to understand the principle of the avc operation in each case. Figure 8-4 of Ch. 8 shows a broadcast receiver with avc.

7-4. Mixers and Converters.[13] The use of the terms "mixer and "converter" is not clean-cut. They are sometimes used interchangeably, sometimes together (as "a converter mixer"), and sometimes to designate distinctly different devices. Following the last usage the term "mixer" will be limited to a tube and associated circuits that combine a modulated input wave with a wave from a separate local oscillator effectively to change the carrier of the original modulating signal from that of the input wave to that of the local oscillator. This process is frequently referred to as detection, although the distinction between detection and modulation is not as clear here as in other cases.

The term "converter" will be used to denote a tube and associated circuits in which the local oscillator and the "detector" or mixer tubes are combined in a single envelope.

The need for the combination of a local oscillator and a mixer, i.e., a converter, arises primarily in superheterodyne receivers, and since the superheterodyne principle is used in most receivers, it is desirable to consider the reasons for utilizing it.

In a superheterodyne receiver the incoming modulated wave, whatever its carrier frequency, is combined (often after one or more stages of amplification) with the wave from an oscillator of such frequency that the useful resultant wave always has a given frequency (intermediate frequency or i-f.) within relatively narrow limits. This permits the use of <u>fixed</u> tuned circuits in the following (i-f) amplifier, and this system can be made to yield improved selectivity, sensitivity and fidelity.

Evidently if the receiver is to respond to incoming signals which may be located anywhere in a relatively broad band of frequencies (for example, to recieve any station in the commercial broadcast band), the local oscillator will have to be

13. In circuit diagrams and in some literature the mixer or converter is often called the first detector.

variable in frequency, so that the difference[14] between the local oscillator frequency and the carrier frequency of the selected incoming wave is equal to the fixed intermediate frequency. (In the receivers for commercial broadcasting, the i-f. is usually in the neighborhood of 460 kc; in receivers for other bands the i-f. may, and probably will, be different; sometimes several i-f stages, each with a different i-f frequency, are used in one receiver.) The maintaining of the local oscillator at the proper frequency leads to the problem of frequency control, closely allied to the use of reactance tubes in frequency modulation (Ch. 6). The appropriate changing of frequency of the local oscillator[15] as the carrier frequency of the incoming signal is changed creates rather difficult problems of "tracking

From the preceding discussion of modulation (Ch. 6) and detection (this chapter) it will be realized that in either the mixer or the converter the plate current will contain many components, some of which will have frequencies that are the sum and the difference of the frequencies of the various sine waves introduced. In the case of an amplitude-modulated wave (eq. 7-2), and a wave from a local oscillator, there will be terms of frequencies ω_i, $\omega_i + \omega_s$, and $\omega_i - \omega_s$ where ω_i is the i-f. and ω_s the angular frequency of the original modulating signal. Although components of many other frequencies are present, appropriate tuning of the output circuit may eliminate them. Hence the original signal of angular frequency ω_s may be made to appea on a carrier of frequency ω_i. Since ω_i is the same whatever the original carrier may be, circuits immediately following the mixe

14. Often, but not necessarily, the oscillator frequency is greater than that of the carrier, and difference frequencies are used. In the broadcast band, 550 to 1600 kc, with i-f. about 460 kc, it is desirable to have the oscillator of higher frequency than the carrier because the range of the oscillator (1010 to 2060 kc) is a much smaller fraction of the lowest frequency (1010 kc) than would be the case if the sum frequencies were used.

For other than broadcast receivers, intermediate frequencies run from one to 100 Mc using standard tubes, and from 100 to 500 Mc using special tubes, and of course may be higher if need arises. When the first i-f. is relatively high, numerous i-f stages may be used, the i-f.'s decreasing from one to the next.

15. Local oscillators in general should have very little noise or other modu lation in their output, and should have a high degree of frequency stabi ity. At u-h-f. the oscillators discussed in Ch. 10 may be used, as well as some not mentioned there because they are not commercially available. Klystrons and other velocity-modulated tubes are reasonably rugged and have relatively large output as local oscillators

Signal-to-noise ratio is extremely important in local oscillators, be cause there is little chance of circumventing noise originating there.

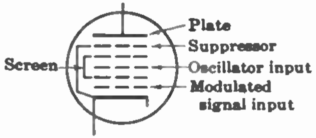
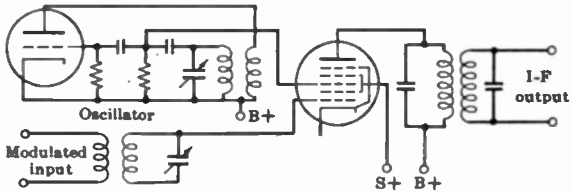
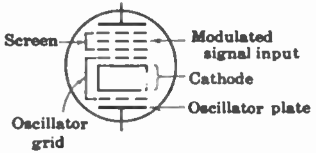
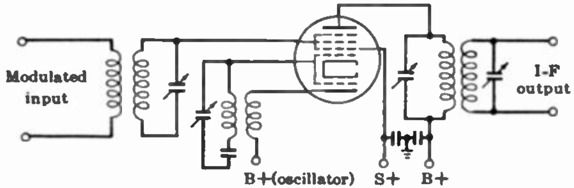
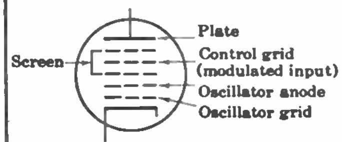
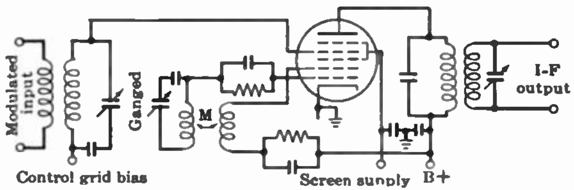
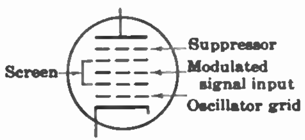
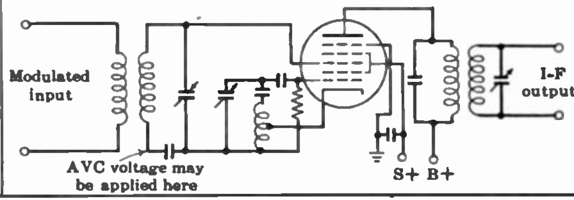
<table>
<tr><th>Electrode Uses</th><th>Possible Circuit</th><th>Name and Comment</th></tr>
<tr><td>
</td><td>
</td><td>(a) Pentagrid mixer and a separate local oscillator.</td></tr>
<tr><td>
</td><td>
</td><td>(b) Triode hexode converter. The triode (bottom) is in the oscillator circuit and feeds directly into the first grid of the hexode mixer (top). The combination forms a converter.</td></tr>
<tr><td>
</td><td>
</td><td>(c) Pentagrid converter, electronic coupling.</td></tr>
<tr><td>
</td><td>
</td><td>(d) Pentagrid converter of type different from (c).</td></tr>
</table>

Fig. 7-8. Examples of Mixers and Converters.

or converter need be designed only to handle ω_i plus or minus the maximum ω_s to be transmitted.

Figure 7-8 shows some tubes and circuits suitable for use as converters and mixers. Each arrangement has some advantages and some disadvantages, most of which are concerned with noise, isolation of various circuits, conversion transconductance, etc.

Conversion transconductance, g_c, is a quantity similar to the usual mutual conductance, g_m, of a triode, or to the control-grid to plate transconductance of a multielectrode tube; it is the ratio of an increment of current in the i-f transformer primary to the corresponding increment in the r-f carrier voltage which produced it. An indication of order of magnitude of conversion transconductance and the variation of this quantity with current flowing in the oscillator grid of a converter is given in Fig. 7-9. The fact that g_c is not constant but definitely varies with grid current is of importance.

7-5. Detection of Frequency-Modulated Waves: the Discriminator.- It was shown in Ch. 6 that an F-M wave can be considered as the sum of carrier frequency and numerous side-frequency components. Assuming a superheterodyne receiver, the

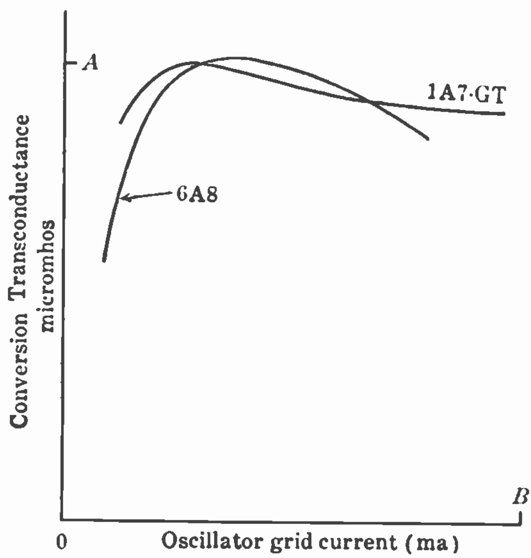


FIG. 7-9. Conversion transconductance, under normal operating conditions, for two pentagrid converters. A = 250 for the 1 A 7, 500 for the 6 A 8; B = 0.1 ma for the 1 A 7, 1 ma for the 6 A 8.

F-M wave can be received on the antenna, amplified in any r-f amplifier stages present, passed through the first detector (converter or mixer), and then through the i-f amplifier, which must have an appropriately wide pass band. It will then appear as an F-M wave with the i-f. for carrier at the output terminals of the i-f stage. Figure 8-6 of Ch. 8 gives a block diagram. The next process is "limiting," that is, removing all amplitude modulation present. This is one of the most essential processes in

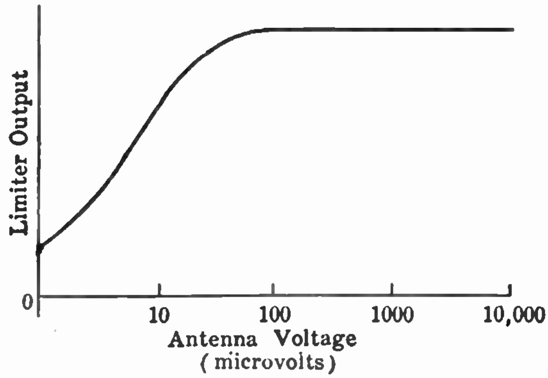


FIG. 7-10. Relative limiter output vs antenna peak voltage The curve illustrates the desirability of having a sufficiently large input signal (greater than 100 μv in the case shown) to have the limiter function properly. (From Everitt, Elec. Eng., 59, No. 11,613, Nov. 1940.)

F-M receivers, and accounts for much of the superiority of an F-M system over an A-M system when such superiority exists. Figure 6-8c shows an F-M wave with A-M, the latter presumably being distortion, noise, etc., picked up anywhere between the original modulation at the transmitter and the output terminals of the last i-f stage of the receiver. The limiter (usually a pentode that clips off the peaks; the plate characteristic of a pentode is nearly straight from the origin to the top section, which is almost flat) changes the curve of Fig. 6-8c to that of Fig. 6-9. Thus virtually all spurious signals that have caused amplitude modulation (some will cause frequency modulation) are eliminated. The fact that the curve of Fig. 6-9 has flat peaks and is not the same curve as would result from a straight frequency-modulated sine wave is not of great importance.

The next step in the detection of an F-M wave is to pass the wave through a circuit that will change the limited wave of Fig. 6-9 to one having both A-M and F-M, but with the A-M representing the same signal as the F-M. This change can be accomplished by impressing the limited F-M wave of Fig. 6-9 on a slightly off-tuned parallel resonant circuit. Roughly, the reader can see this action by picturing the response of such a circuit as the frequency varies back and forth during its modulation period.

The combined amplitude- and frequency-modulated wave is then supplied to a detector which responds to the envelope of the wave and ignores for practical purposes the frequency modulation that remains in the wave. To reduce distortion in the resultant demodulated wave, a balanced detector is usually used. The combination of tuned circuit and detector is called a discriminator. Figure 8-6 gives a detail of a limiter and discriminator circuit.

The determination of the response of simple circuits to F-M waves is not simple, and it does not appear advisable to go into the subject in greater detail here. For an introduction to the theory in general form, reference may be made to the two papers listed in the footnote.[16]

7-6. Detection at U-H-F.- Detection at u-h-f. is the same in principle as at that lower frequencies. Two differences may be noted: (1) the band-width to be handled is often quite large, which tends to decrease detector efficiency; and (2) crystal detectors are used far more extensively at u-h-f. than at the lower frequencies.

The term "crystal detector" refers not to a quartz or other similar crystal exhibiting a piezoelectric effect (such as is used for frequency control, etc.), but to a combination of a crystalline material such as silicon and a fine point contact of metal such as tungsten that possesses a non-linear current-voltage characteristic. A current-voltage curve for a crystal detector of the type here described is somewhat similar to the diode characteristic shown in Fig. 7-3b. Consequently a crystal detector may often be used in place of a diode in modulation, detection, rectification, and in other functions.

Crystal detectors have definite limitations, and they do not by any means displace the special diodes available for u-h-f use. As ordinarily employed, with a fine wire or "cat whisker" in contact with a conducting or semi-conducting crystal, the power that can be handled is very small, temperature changes may affect the crystal, and the resistance to current flow in "the back direction" is frequently not very large. Capacitances

16. Carson, John R., and Fry, Thornton C., Variable Frequency Electric Circuit Theory with Application to the Theory of Frequency Modulation, Bell Sys. Tech. Jour., 16, No. 4, 513, Oct. (1937).

Roder, Hans, Effects of Tuned Circuits upon a Frequency Modulated Signal, Proc. I.R.E., 25, No. 12, Dec. (1937).

between parts of the crystal holder have shunting effects, but the crystal impedance is usually not greater than several thousand ohms. On the other hand, a crystal is an easily manipulated device, not difficult to make, and it can be manufactured in large quantity at relatively low cost. Crystals used as mixers appear on the whole to have the highest signal-to-noise ratio that can be obtained in some frequency regions, and they are often used for this purpose at frequencies above 300 Mc.

Figure 7-11 shows the form of typical crystal characteristics, and Fig. 7-12 gives curves of backward and forward resistance of a copper copper-oxide rectifier which is similar to

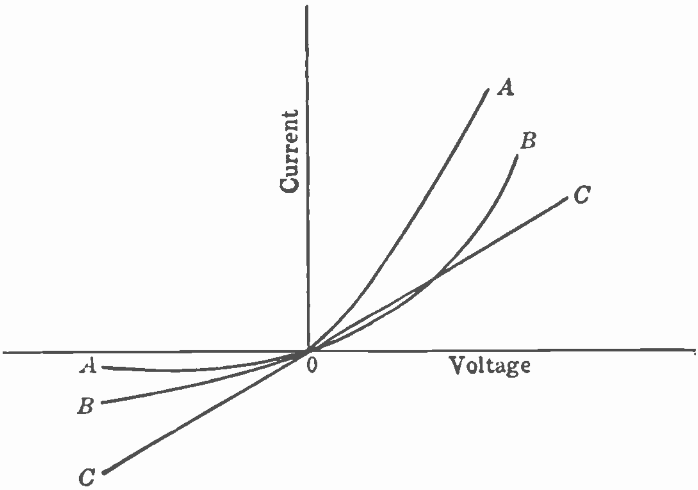


FIG. 7-11 General form of crystal characteristics. Curve A is roughly linear to the right of the origin O, B approximates a square law in this region, C is the type of curve obtained for contact at a point on the crystal which exhibits no directional properties.

the crystals discussed here, except that the contact between crystal and "whisker" is not small, as it is in a detector comprising a fine tungsten wire touching a silicon crystal, but is relatively large because of the fact that the copper oxide is formed on the copper over a relatively large surface. (The reader may consider copper the "crystal" and copper oxide the equivalent of the "whisker" multiplied many times.) Copper copper-oxide rectifiers are used extensively at the lower frequencies as modulators and rectifiers. For the crystals used at u-h-f., the ordinates of the ratio curve will be much smaller than indicated in Fig. 7-12c.

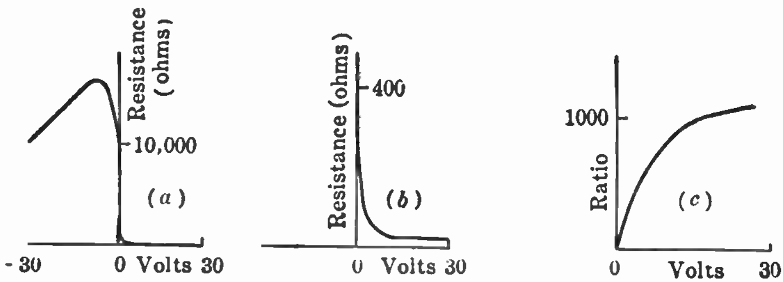


FIG. 7-12. Copper copper-oxide rectifier curves; (*a*) resistance vs applied voltage The voltage is considered positive when relatively large currents flow; (*b*) is an enlargement of the curve of (*a*); (*c*) gives the ratio of current in one direction to current in the opposite for a given voltage (From Grondahl, Rev. of Mod. Phys.,- V 5-141, April 1933.).

Chapter 8

RADIO RECEIVERS

A radio receiver is such a common device that it needs no definition regarding its function in radio communications. There are, however, several important considerations that must be recognized in the design of a good radio receiver. A high-quality radio receiver will faithfully reproduce the intelligence, speech, music, or other signal that was originally impressed on the modulated wave at the transmitter. The reproduction of the intelligence or signal should not be accompanied by excessive noise, interference from unwanted radio transmissions, or distortion. An ideal system for radio transmission and reception would also be completely free from disturbances from atmospherics but, practically, a certain amount of response to these disturbances must be expected because this is inherent in any system.

8-1. Types of Noise.- Radio receivers generate a certain amount of noise disturbance that does not appear in the signal picked up by the antenna. The effect of these noise voltages depend not upon their absolute magnitude but upon the ratio of the output power from the signal to the output power from the noise. In an absolute sense, noise disturbances in a radio receiver increase as the amplification of the receiver is increased. Weak signal voltages require high amplification in the receiver. Consequently, the problem of reducing unwanted noise disturbances increases in difficulty with increase in the sensitivity of receiver. The sensitivity of a radio receiver is that characteristic which determines the minimum strength of signal input capable of causing a desired value of signal output.[1] There are four types of disturbances[2] in a radio receiver that are classed as noise. These are thermal agitation, shot effect, microphonics, and hum from a-c operation. These will be defined and described in turn.

Thermal Agitation. A coordinated flow of electrons through a conductor may be caused by establishing a difference of potential between the terminals of the conductor. Such a coordinated flow of free electrons consitutes the familiar electric conduction current. Superimposed upon this steady flow of

1. Definition taken from "Standards on Radio Receivers," I.R.E., 1938.
2. Note that the discussion refers to noise created in the receiver, and does not consider noise received by the receiver.

electrons are minute variations that change with time in an irregular fashion caused by the fact that the free electrons of a conductor are also in random motion, with velocities that depend upon the temperature. The resultant motion of the electrons in a conductor at any particular instant of time produces a fluctuating voltage between its two terminals. "Thermal agitation" is the name given to the continual uncoordinated movement of free electrons that goes on inside a conductor. Since the fluctuating voltage set up by thermal agitation cannot be represented by a combination of components of constant amplitude and frequency even for a very short interval of time, it is generally represented as a root-mean-square value obtained by integration of all components between specified limits of frequency.

It has been found that the noise voltage generated in a frequency band of given width is potentially independent of the location of the band in the frequency spectrum. The thermal agitation noise voltage is given by the equation[3]

$$E^2 = 4\,kT \int_{f_1}^{f_2} R\,df \qquad (8\text{-}1)$$

where E is the root-mean-square noise voltage

- k is Boltzman's constant 1.37×10^{-28} joules per degree absolute.
- T is the absolute temperature in degrees Kelvin.
- R is the resistance in ohms.
- f is the frequency in cycles per second; f_1 and f_2 are limits of the frequency band.

When the above integration is carried out over a band of frequencies Δf wide, for which the resistance is constant, the expression results

$$E = \sqrt{4\,k\,T\,R\,\Delta f} \qquad (8\text{-}3)$$

Thus for any resistor R having a temperature T there is an equivalent noise voltage of magnitude E given by (8-3). If this noise voltage E is developed across a resistor located in the input end of a receiver it will be amplified along with the signal. Since for successful reception there is a definite minimum ratio of signal voltage to noise voltage, one practical factor limiting the sensitivity of a receiver is the extent to which the voltage from thermal agitation has been reduced in the input stage or stages.

Shot Effect. The shot effect in a radio receiver or any vacuum-tube amplifier is caused by the discrete particle nature

3. Johnson, J.B., Phys. Rev., 32, 97 (July 1928); Jansky, K.G., "An Experimental Investigation of the Characteristics of Certain Types of Noise," Proc. I.R.E., Vol. 27, No. 12 (December, 1939).

of the flow of electrons from the cathode to the anode of a tube. The spasmodic flow of the discrete particles causes a minute current of irregular nature to be superimposed upon the main plate current of the tube. The current caused by shot effect sets up a minute voltage across the output impedance. This voltage is then amplified, along with the signal voltage, to the output of the receiver. Shot effect is most pronounced when a tube is operated with very little space charge around the cathode. Consequently the effect can be minimized by operating at such electrode voltages and cathode temperature to insure a generous space charge around the cathode at all times. Like thermal agitation effect, the noise energies from shot effect are spread over the entire frequency spectrum.

Microphonic Noise. Microphonic noise is caused by mechanical vibration of the elements of the tubes. The mechanical vibration may be due either to mechanical vibration that travels through the chassis and tube sockets or to acoutical vibrations striking the tubes directly. The direct transmission of mechanical vibration can be prevented by supporting the tube sockets on springs or other resilient material, or by shock mounting the entire chassis as is always done in aircraft radio equipment. Acoustical vibration can be kept away from the tubes by enclosing them individually, or the set as a whole, in a case through which sound waves do not readily pass.

Hum. Hum voltages appearing in the output of a receiver are usually caused by operating the receiver from an a-c source. Hum frequencies are multiples of the a-c source frequency. Again, the effect of hum depends on voltage or current ratios and not on absolute magnitudes. In order to reduce the effect of hum, care must be taken in the input of low-level stages of the receiver, particularly the first audio stage. Hum voltages may arise from operating the filament or heater on a-c, from insufficient filtering in the plate source when using a rectifier, from improperly filtered cathode bias resistor, and from stray magnetic and electric fields inducing voltages in the various circuit components.

Hum is usually "picked up" by or originated in the audio-frequency stages of a radio receiver where the pass band of the amplifier includes the hum frequencies. However, hum can also originate in the radio- and intermediate-frequency stages where because of a certain amount of non-linearity between output and input voltages the hum voltage superimposes modulation of the hum frequency upon the regular modulated wave. Very serious sources of hum are found in audio-frequency transformers operating at low signal level unless the transformers are carefully shielded and so wound as to reduce to a negligible extent the effects of external magnetic fields. Hum voltages induced by electric fields of the power supply and by other stray fields can be eliminated with comparative ease by locating electrostatic shields around the parts most susceptable to such disturbances.

8-2. Characteristics of a Radio Receiver.- The specifications of a radio receiver that relate to the quality of its performance are usually expressed in the following terms: selectivity, sensitivity, and fidelity. An additional specification relates to the noise output or signal-to-noise ratio when the modulated signal input is noise free. Any noise that appears in the output of the receiver under this condition is generated within the set and reduces its quality of performance. The standard tests for selectivity, sensitivity, and fidelity will not reveal the noise output of the receiver. The amount of such noises or disturbances that get through to the output of the receiver are functions of the system of transmission, the selectivity, and the sensitivity of the receiver.

Selectivity. The selectivity of a radio receiver is that characteristic which determines the extent to which it is capable of differentiating between the desired signal and disturbances of other frequencies.[4] Selectivity is generally expressed as a ratio of the r-f input voltages required off resonance and at resonance to produce a standard output voltage or power in a receiver. Generally the selectivity refers to the entire receiver and the input voltages are modulated with a constant modulation factor and frequency for making selectivity tests. Thus selectivity becomes a curve of input voltage ratios versus kilocycles off resonance. A typical curve is shown in Fig. 8-1. In many current broadcast receivers, the selectivity curve might be nearly constant over a band of frequencies of about 10 kilocycles width centered about the carrier frequency.

Sensitivity. The sensitivity of a radio receiver was defined in Sec. 8-1. Sensitivity is usually expressed either in decibels below one volt or directly in microvolts introduced in the input of the receiver through a standard dummy antenna to produce normal test output when all adjustments and controls are set for maximum sensitivity.

The input signal for a sensitivity measurement is a carrier modulated 30 per cent at a frequency of 400 cycles. Normal test output depends somewhat upon the power output of the receiver. For example, for receivers of less than 1 watt and greater than 0.1 watt output, the normal test output is 0.05 watt in the particular load resistor for which the final power amplifier stage of the receiver is designed to operate. The sensitivity of a radio receiver usually varies somewhat over any

4. Definition taken from "Standards on Radio Receivers," I.R.E., 1938.

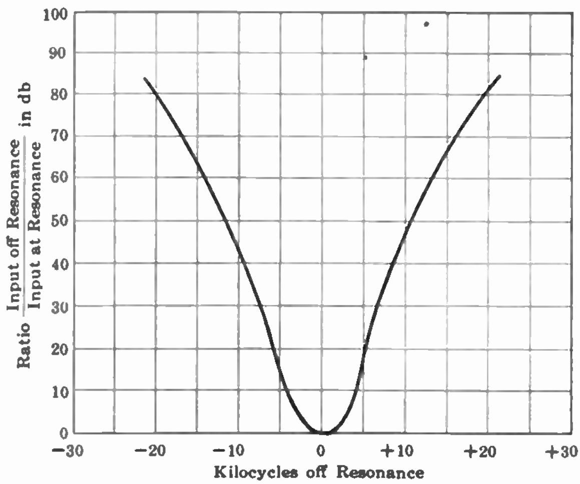


FIG. 8-1. Typical selectivity curve for a radio receiver.

one tuning band and must therefore be given as a curve plotted against frequency.

Fidelity. Fidelity is the degree to which a radio receiver accurately reproduces at its output terminals the essential characteristics of the signal which is impressed on its input. This is essentially the Institute of Radio Engineers' definition. The fidelity of a receiver with sound output is usually separated into electric fidelity and acoustic fidelity. Fidelity measurements are made by keeping the carrier frequency and voltage constant and varying the modulating frequency while the modulation is maintained at 30 per cent. For electric fidelity the receiver is terminated in its normal load and the volume control is adjusted for normal test output. The relative response of the receiver in decibels from an arbitrary zero level at 400 cycles is then plotted against modulation frequency. The Institute of Radio Engineers[4] has adopted some special standard instructions for tuning a receiver that has automatic volume control prior to making fidelity tests.

Noise and Hum. The specifications of a radio receiver should include the noise and hum voltage outputs that will be tolerated when a specified noise-free carrier voltage is applied through a dummy antenna to the input terminals of the receiver.

8-3. Tuned Radio-Frequency Receiver.- Radio receivers may be classifed according to the method by which the modulated voltage at the input is amplified to sufficient amplitude to operate the final detector that converts the signal back to its original unmodulated form. The two types in common use at the present time are the tuned-radio-frequency amplifier and the superheterodyne.

In the tuned-radio-frequency receiver the modulated signal input voltage is amplified through a cascade system of tubes with tuned transformers for coupling circuits. These tuned transformer-coupling circuits are either of the tuned-secondary type or the doubly-tuned type, which were described and treated in Ch. 3. The tuned-secondary type has been used frequently because of the difficulty of getting condensers to track properly

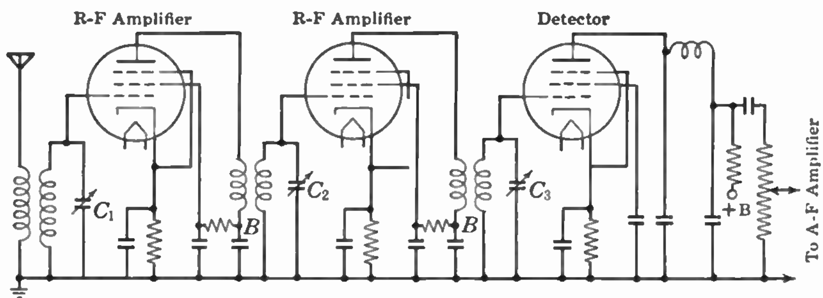


FIG. 8-2. Tuned radio-frequency receiver. R-f. stages and detector only. Condensers C_1, C_2, and C_3 ganged.

in the doubly-tuned type. The tubes are r-f pentodes. A schematic circuit diagram is shown in Fig. 8-2.

The tuned-radio-frequency receiver may readily have a fairly high sensitivity and a comparatively low noise level compared with receivers of other types. It is also free from such things as image and harmonic frequency response that sometimes exist in a superheterodyne receiver. However, the selectivity of a tuned-radio-frequency receiver can not generally be adjusted satisfactorily to reject interfering frequencies outside of the band of desired reception and still have constant gain over this band without requiring complicated tuning adjustments. This is particularly true of a receiver for the standard broadcast frequencies, where the width of the pass band is approximately 10 kilocycles and the range of frequencies is 550 kc to 1600 kc. Consequently, tuned radio-frequency receivers have nearly disappeared from the radio receiver field.

8-4. The Superheterodyne Receiver.- The superheterodyne receiver comprises a cascade arrangement of a radio-frequency amplifier or a pre-selector, a first dectector and its associated oscillator, an intermediate-frequency amplifier, a second detector and an audio-frequency amplifier. A schematic block diagram is shown in Fig. 8-3. The radio-frequency amplification or the preselection is provided in some receivers by a simple tuned transformer, but in better receivers it is obtained by one or more stages of radio-frequency amplification. The function of the

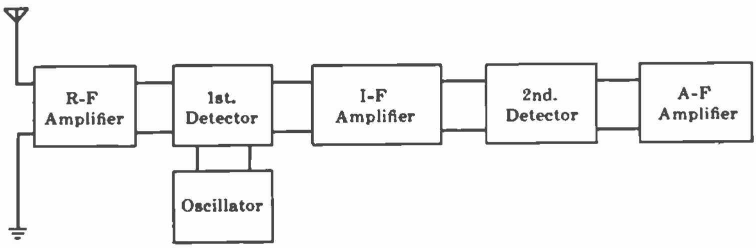


FIG. 8-3. Block diagram of a superheterodyne receiver.

first detector and its associated oscillator is to effect a frequency conversion of the signal from its original location in the spectrum to that of the i-f amplifier. The intermediate-frequency amplifier comprises one or more stages of selective amplification operating at a frequency below the lowest radio frequency to which the receiver can be tuned. The second detector demodulates the modulated signal of i-f frequency and thereby generates voltages that have the same frequencies as the original modulating frequencies at the transmitter. In a broadcast receiver these voltages lie in the audio-frequency spectrum and generally are so weak that they have tc be amplified by one or more stages of audio-frequency amplification to be of sufficient magnitude to drive the final audio-frequency power amplifier.

The First Detector and Oscillator. The function of the first detector and its local oscillator is that of a frequency converter. The incoming signal and a voltage from the local oscillator are supplied to a tube operated on a non-linear part of its characteristic and thereby generate among other frequencies a new modulated signal that is a facsimile of the original except that it has a carrier frequency equal to the difference between the original carrier frequency and the oscillator frequency. This new signal is modulated by the same percentage as that of the original signal. There are a number of circuits for doing this. The circuit shown in Fig. 8-4 is known as a pentagrid converter. The detector and oscillator are contained in the same tube structure. The oscillator section makes use of the first two grids, the second grid becoming the anode of the oscillator. This gives electron coupling to the detector section. Another method somewhat similar to the above is to have

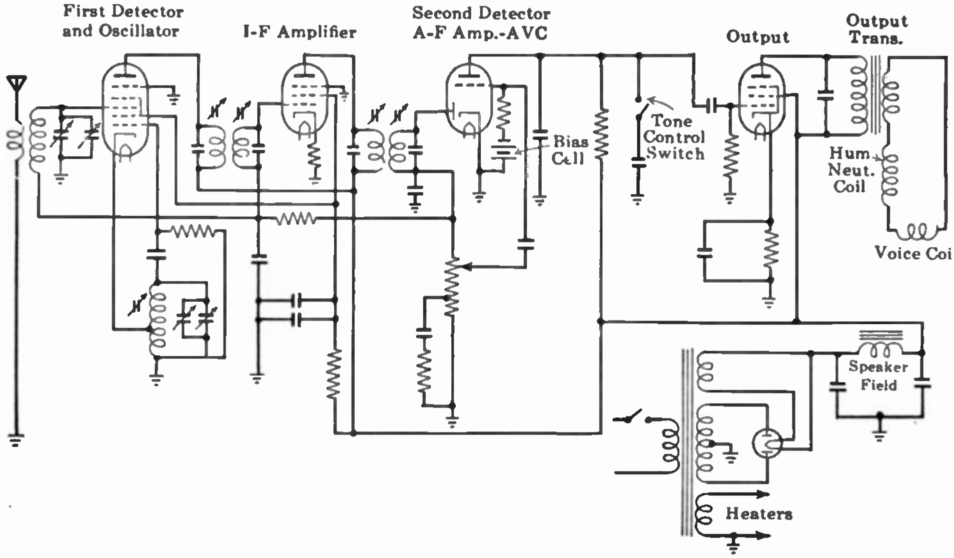


FIG. 8-4. A typical superheterodyne receiver (RCA Dynamic Demonstrator II).

a separate tube for the oscillator and to couple the oscillator to the third grid of a pentagrid mixer type of detector. This latter method is more satisfactory at the higher frequencies where the percentage difference between the original signal and oscillator frequency is small.

The Radio-frequency Amplifier. The radio-frequency amplifier of a superheterodyne receiver must be sufficiently selective to maintain a fairly high ratio of desired signal to unwanted signals on the first detector. The extent to which this is carried out in practice depends largely on the cost of the receiver. Usually, however, not more than one stage of radio-frequency amplification is employed. The selectivity performance of a receiver that has no r-f amplification cannot be expected to equal that of a receiver that has one stage of r-f amplification. This is especially true if the receiver is operated under conditions where there are two strong signals of comparable magnitudes and separated in frequency by twice the intermediate frequency of the receiver. When the set is tuned to the signal of lower frequency and the oscillator frequency is above the signal frequency there will be a disturbing response from the signal of higher frequency unless the r-f amplifier has greatly increased the ratio of the two signals at the first detector tube. This is known as image-frequency response. The practice usually followed in "all-wave receivers" at the present time makes use of an intermediate frequency of about 460 kc with the oscillator frequency located above the signal frequency. Thus in a receiver that has a 460-kc intermediate-frequency amplifier, image-frequency response is possible from an interfering signal that reaches the first detector and has a frequency 920 kc above the desired signal.

The Intermediate-frequency Amplifier. The intermediate-frequency amplifier of a superheterodyne receiver consists of one or more stages, generally of the doubly-tuned transformer-type coupling. These transformers are adjusted to have coupling equal to or a little greater than the critical value and so that they pass both side bands of the modulated voltage. For usual broadcast service this requires a band width of 10,000 cycles centered about the carrier. The tubes of the intermediate-frequency amplifier are conventional r-f pentodes or super-control pentodes.

The Second Detector and Automatic Volume Control. The second detector of a superheterodyne receiver is usually the diode type. For the receiver shown in Fig. 8-4 the second detector and automatic volume control are incorporated in one tube. This is a customary practice although in some cases a second diode is used for the automatic volume control. The d-c current generated by the diode is proportional to the r-f signal voltage. This d-c current establishes the bias voltage for the first and second tubes. These tubes are the super-control type for which the output signal voltage is an inverse function of the bias voltage. When the strength of the signal to the receiver changes, the bias voltage changes. This in turn causes the sensitivity of the receiver to change in such a way as to keep a nearly constant signal on the second detector. There must be some change in the output in order to have automatic control action. Delayed automatic volume control means that the action of the control is delayed until the signal to the second detector reaches a certain value.

Manual Volume Control. In most receivers that have automatic volume control, the manual volume control is effected by a potentiometer in the diode detector circuit across which the audio signal voltage is developed. The movable arm of the potentiometer is connected to the grid of the first audio-frequency amplifier tube. In some receivers that have no automatic volume control, the manual volume control changes the bias to one or more super-control tubes in the r-f and i-f stages. This control is also sometimes arranged to perform the dual function of changing the signal input to the receiver as well as changing the sensitivity of the receiver.

Automatic Frequency Control. Automatic frequency control in a superheterodyne receiver refers to circuits for the maintenance of the frequency of the oscillator at a constant value. It also provides an automatic tuning of the oscillator to the correct frequency for producing the correct intermediate frequency when the oscillator tuning is near the correct adjustment or in cases where either carrier or local oscillator frequency drifts. This feature is desirable in a receiver that has a high degree of automatic volume control because in such a receiver it is difficult to tell just when the tuning is correct. Figure 8-5 illustrates a method for obtaining automatic frequency control. The duodiode and its associated circuit is called a discriminator. It generates a d-c voltage across its load resistor that depends upon the ratio of the impressed frequency to

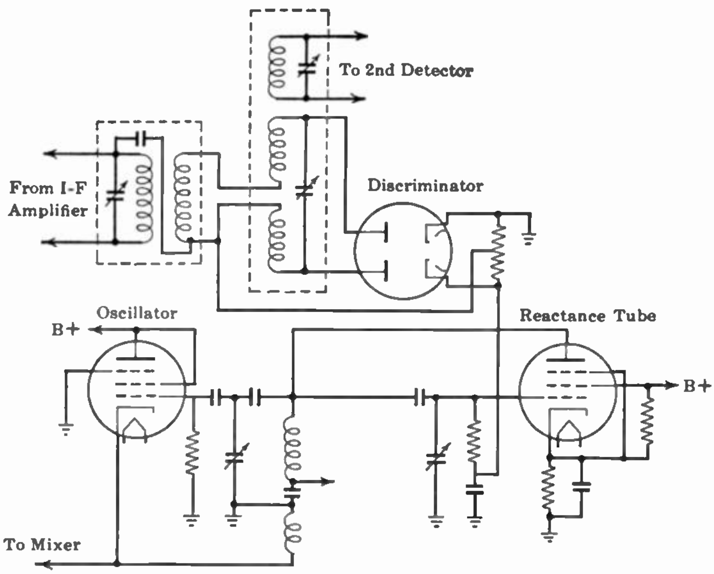


FIG. 8-5. Typical circuit diagram of an automatic frequency control for a radio receiver.

the resonant frequency of the circuit connected to its plates. Use is made of phase shift that occurs in the neighborhood of the resonant frequency in a resonant circuit. The duodiode and its associated resonant circuit is so arranged with respect to the intermediate-frequency amplifier that a change in frequency causes one diode to receive a greater voltage than the other. This results in a net d-c voltage across the combined load resistors. This d-c voltage is used to bias a tube sometimes known as a variable reactance tube. The variable reactance tube[5] is

5. This is the same type of tube as is used for frequency-modulating a carrier. See Ch. 6.

connected to the oscillator so that the r-f voltage applied to the grid is 90° out of phase with that applied to the plate. The plate current of the variable reactance tube is 90° out of phase with the plate potential and thereby becomes a part of the tuning reactance of the oscillator. The magnitudes of the plate current and the reactance depend upon the bias of the tube; this in turn depends upon the frequency shift at the discriminator. Hence, if the intermediate frequency is slightly off value or changes momentarily, the oscillator is so tuned by the discriminator and variable reactance tube that the intermediate frequency is immediately brought back in tune with the amplifier.

Tone Control. Most radio receivers are equipped with a tone control. The simpler type of tone control attenuates only the higher frequencies. One method for doing this is to connect a fixed condenser and variable resistor in series from the plate of the first amplifier tube to ground, or cathode. As the value of the resistor is reduced, the higher frequencies are attenuated because the gain of the tube at the higher frequencies is made less.

Cross Modulation. Cross modulation, sometimes known as crosstalk, in a radio receiver is the name given to unwanted signals in the output which are caused by the interaction of one or more undesired r-f signals with the r-f signal for which the receiver is tuned. One such type of crosstalk exists only in a superheterodyne receiver. This type, as was pointed out earlier, is due to image-frequency response. Another type can exist in any type of receiver when the receiver is tuned to a strong signal and the sensitivity of the receiver is reduced by operating one or more amplifier tubes near cutoff bias. Then because of the high order of curvature of the tube characteristic near cutoff, a second simultaneous signal of a different frequency will modulate the desired signal and cause unwanted output. For this kind of crosstalk there is no specific relation between the frequencies of the desired and undesired signals. The magnitude of the crosstalk depends upon the curvature of the amplifier tube characteristic and the strength of the undesired signal that reaches the grid of the tube. This type of cross modulation may be reduced by attenuating the undesired signal as much as possible before it reaches the first control tube, and it may be avoided by using a tube that does not have a high order of curvature over any part of its characteristic. The type of tube that has this characteristic is known as the super-control, or variable-mu, tube. With the use of these two expedients crosstalk has been largely eliminated in present-day receivers.

Distortion. Distortion in a radio receiver is due to a number of things. The detector causes some distortion which increases with the percentage of modulation. This is a type of wave-form distortion. The r-f and i-f amplifiers may cause distortion because the gain of these amplifiers is not constant over the full range of the side-band frequencies. This latter type of distortion is frequency distortion because some of the higher side-band frequencies are not amplified at the same rate as the lower frequencies. It can be reduced by broadening the pass band of the amplifiers, but generally this can only be done at the expense of sensitivity. The audio-frequency amplifier may also cause both wave-form and frequency distortion. In general, the distortion caused by the audio-frequency amplifier can be reduced only at a sacrifice of gain. The loud speaker for a radio receiver also causes considerable distortion because it does not have constant sound output for all audio-frequency signals of the same amplitude. In summary it can be said, in general, that a radio receiver does not faithfully reproduce the modulating signal impressed on the input and that high fidelity and high sensitivity are incompatible.

8-5. Receiver for a Frequency-Modulated Signal.- A typical receiver for a frequency-modulated signal is of the superheterodyne type. It is quite similar in circuit layout to an amplitude-modulation receiver up to the stage where the modulating signals are recovered from the modulated wave. Figure 8-6 shows a block diagram of such a receiver and a detailed diagram of the demodulator, or discriminator, with the limiter tube. Up to the limiter tube there is little essential difference

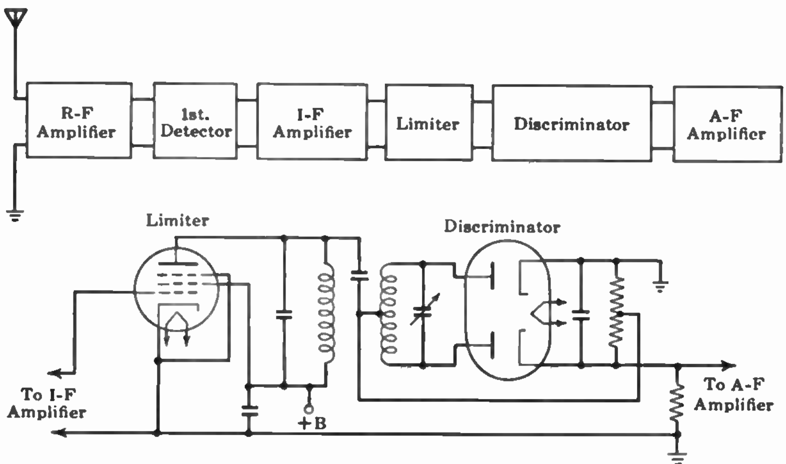


FIG. 8-6. Block diagram of an F-M receiver with detail on limiter and discriminator.

between the F-M (frequency-modulation) and the A-M (amplitude-modulation) receivers. As a matter of fact, receivers are built for ultra-high-frequency service that will operate on either F-M or A-M and a large part of the receiver is used as a common section for either service. For A-M service, the conventional A-M demodulator is switched into the circuit between the output of the intermediate-frequency amplifier and the audio amplifier. For F-M service, a limiter tube and a discriminator replace the A-M demodulator. All-wave combination receivers generally have separate intermediate-frequency transformers for A-M and F-M service, because the intermediate frequency of an F-M receiver has to be higher than that of the conventional A-M receiver. The maximum band width that has been adopted as standard for an F-M signal is 200 kilocycles. To amplify a band this wide without undue attenuation of the frequencies near the edges of the band requires a much higher intermediate frequency than is necessary in an A-M receiver. Intermediate frequencies from 4 to about 6 megacycles are being used.

The Limiter and the Discriminator. Because the limiter and discriminator are the only essential parts of an F-M receiver that are different from and A-M receiver, only a brief description of these will be given. As a frequency-modulated wave is one of constant amplitude it is possible to remove any amplitude modulation from it without detracting from the quality of the demodulated output. Hence, any amplitude modulation that the frequency-modulated wave may have acquired by the time it reaches the output of the intermediate-frequency amplifier can be eliminated. This is done by the limiter tube. The limiter tube operates with low plate potential and little bias, and is driven to plate current saturation. Therefore, any additional increase in amplitude due to amplitude modulation produces no further increase in output voltage.

From the output of the limiter tube the F-M signal goes to the discriminator. The discriminator performs in the same manner as does the discriminator in the automatic frequency control of an A-M receiver, except that it produces an audio-frequency output voltage because the signal input voltage is changing in frequency about the carrier. The amplitude of the audio output is proportional to the instantaneous signal frequency deviation from the carrier frequency, and the audio frequency is proportional to the rate at which the signal frequency is deviating.

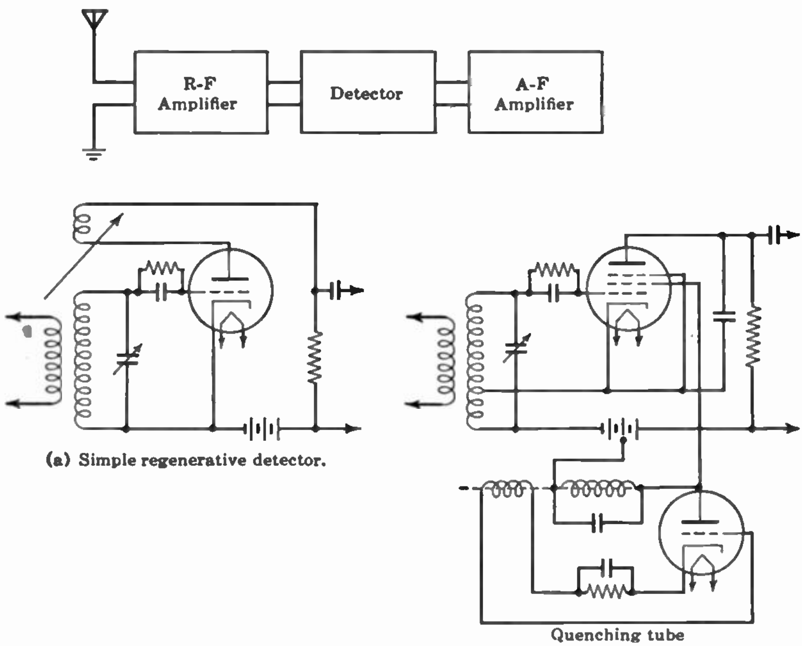
8-6. The Pass-Band of a Radio Receiver.- What has been said so far about radio receivers has been largely from the standpoint of the use of such receivers for voice- and music-modulated carrier waves. For voice, or music, reception of amplitude-modulated signals the audio-band width necessary for reasonably good quality is in the neighborhood of 10,000 cps centered about the carrier. This means that frequencies between some small number and 5000 cps are passed, others eliminated. For frequency-modulated signals the present practice is to do better by allowing about 15,000 cps band width. For services other than voice or music transmission, the pass-band width of the receiver depends upon the type of service. For telegraph service, the pass-band width need be only 100 to 300 cps and depends upon the number of letters per minute. A quartz crystal can be used in a telegraph receiver as the coupling element between the tubes of the intermediate-frequency amplifier, because of the narrow band required. This greatly increases the selectivity. A television receiver requires for the picture channel the widest pass band of all. The intermediate frequency of such a receiver must be high: 12.75 kc has been standardized. The intermediate-frequency amplifier must pass a band of frequencies of about 4 megacycles and must have practically constant velocity of transmission. So, also, the amplifier following the second detector must pass about 4 megacycles with nearly constant velocity of transmission.

A receiver for any special service such as periodic impulses or non-recurring and semi-transient impulses must pass a band width equal to the difference between the highest and lowest frequencies contained in, or necessary to form, the impulse. For such a receiver it is essential that the velocity of transmission[6] be extremely constant over the pass band, or the shape of the received impulse will be different than the original transmitted impulse even though the amplitude response of the receiver is constant over the pass band.

8-7. Other Types of Receivers.- Regenerative and Superregenerative Receivers. A regenerative receiver is one in which a regenerative detector is used. Figure 8-7a shows a simple

6. By constant velocity of transmission is meant that change in phase angle is directly proportional to frequency. The common practice of talking about velocity of transmission in a circuit with lumped parameters (R, L, C, tubes, etc.) comes from analogy with transmission line theory, where $\beta\ell = 2\pi\ell/\lambda$ is the change in phase of the incident wave between input and output. Since in the same theory $c = \beta/\omega$ where c is the velocity of phase propagation (see Ch. 11), it is seen that if β (and hence $\beta\ell$) were proportional to ω, c would be constant. Thus, constant velocity corresponds to direct proportionality between phase change and ω. The need, from the steady-state viewpoint, of the latter for distortionless transmission is shown in Ch. 1.

regenerative detector. In this type of detector the plate circuit is coupled into the tuned grid circuit for the purpose of building up the input signal. This is a form of regenerative feedback where the feed-back voltage is in phase with the input voltage. Thus the gain of the tube is greatly increased because with this type of feedback the gain becomes equal to the gain without feedback divided by $1-A\beta$, where[7] $A\beta$ has a positive real term around the resonant frequency of the input circuit. Enormous sensitivity can be obtained with a single tube. There are, however, objections to this type of detector. High regeneration is obtained at the expense of instability. That is, for high sensitivity the adjustments must approach the critical adjustment which, if slightly exceeded by a change in the d-c operating potential or a slight over-adjustment, will result in sustained oscillations. Thus such a detector requires delicate adjustment of the regeneration to keep it functioning properly. Furthermore, when the regeneration is adjusted for high



(a) Simple regenerative detector.

(b) Superregenerative detector.

FIG. 8-7. Regenerative and superregenerative receivers.

7. A is the magnitude of the voltage amplification without feedback and β is the magnitude of the feed-back ratio (see Ch. 3).

sensitivity, the selectivity becomes so high that only a very narrow band of frequencies will be passed.

The superregenerative system of Fig. 8-7b is employed at rather high frequencies. It is used to some extent in portable radio receivers for police communications because of its relative freedom from ignition interference. The sensitivity of such a system is extremely high. The audio output does not depend much on the strength of the input signal. That is, there is a kind of limiter action that causes strong signals to give little more output than weak signals. This type of detector makes use of the high gain that can be obtained in a regenerative detector when adjustments are on the threshold of oscillations. Because of the signal, oscillation starts but the quenching tube which operates at a much lower frequency prevents the oscillations from building up to a very high amplitude before they are stopped.

Communications Receivers. A receiver that has a heterodyne detector is required for receiving telegraph code. Hence, almost any receiver can be converted into a code receiver by the incorporation of a beating oscillator that supplies to the detector a voltage in addition to the signal voltage. The detector can be practically any type used for amplitude modulation. Ordinarily the beating oscillator frequency differs from the signal carrier frequency by about 1000 cycles, although in most code receivers this frequency can be adjusted to suit the listener. Superheterodyne receivers built especially for code reception generally have quartz-crystal coupling in the intermediate-frequency amplifier for limiting the pass band. The beating frequency is introduced into the input of the second detector from a single tube oscillator.

READING LIST

Eastman, A. V., "Fundamentals of Vacuum Tubes," McGraw-Hill Book Co., Inc.

Glasgow, R.S., "Principles of Radio Engineering," McGraw-Hill Book Co., Inc.

Henney, K., "Radio Engineering Handbook," McGraw-Hill Book Co., Inc.

Institute of Radio Engineers, "Standards on Radio Receivers, 1938."

Smith, F.L., "Radiatron Designer's Handbook," R.C.A. Manufacturing Co., Harrison, N.J.

Terman, F.E., "Radio Engineering," McGraw-Hill Book Co., Inc.

Chapter 9

TRANSMITTERS

A radio transmitter is essentially an assembly of radio equipment such as tubes, condensers, resistors, inductors, d-c power supplies and the like for producing power at radio frequencies that is delivered to a radiator and is controlled or modulated by the intelligence to be transmitted. Most of the components of a transmitter, including the frequency controlling device (usually a quartz-crystal oscillator), the d-c power supply, and the various amplifying stages, which may be considered as applications of r-f amplifiers, have been treated separately in other chapters of this text. The transmitter is simply a comprehensive co-operative assembly of such components that will operate at the desired frequency with the desired power output and with the kind of modulation necessary for the service for which it is intended. The components are usually assembled in a steel rack with panels containing the various switches, controls, and meters that are necessary for adjusting the transmitter and monitoring its performance.

9-1. Classification and Frequency Tolerance of Transmitters.- Radio transmitters are classified according to the type of service for which they are applied. There are several classifications as follows:

1. Telegraph code transmitter.
2. Radio-telephone amplitude-modulated transmitter.
3. Radio-telephone frequency-modulated transmitter.
4. Impulse-modulated transmitter. This type is not essentially unlike the radio telephone. A television transmitter is an example of this type.
5. Tone-modulated transmitter. This type is about the same as a radio-telephone transmitter.

Each of these classifications of transmitters may be subdivided according to the particular service and frequency for which the transmitter is used. The design problems connected with the layout and assembly of a particular type depend to some extent upon the operating frequency. Hence there are such classifications as amateur, standard broadcast, ultra-high-frequency broadcast, police, low- and high-frequency telegraph, aircraft mobile, and others.

The frequency tolerances of transmitters are regulated in this country by the Federal Communications Commission. These tolerances have been reduced from time to time because development has made such reductions possible. In some instances, a very high degree of frequency stability is necessary for technical reasons alone. Consequently, any table of tolerances given in this text should not be regarded as final, and anyone interested at any time in the latest figures should write to the Federal Communications Commission, Washington, D.C., for the current tolerance regulations. The present tolerances are

Frequency Band and Type of Service	Tolerance
1. 10 to 550 kc	
a. Fixed land, mobile, other than under (b)	0.1%
b. Mobile station between 110-160 kc and 365-515 kc	0.3%
2. 550-1500 kc (standard broadcast)	20 cycles
3. 1500-6000 kc	
a. Fixed station	0.01%
b. Land station	0.02%
c. Mobile station	
1500-4000 kc	
4115-4165 kc	0.05%
5500-5550 kc	
4000-6000 kc	0.02%
d. Aircraft	0.025%
e. Broadcasting	0.005%
4. 6000-30,000 kc	
a. Fixed	0.01%
b. Land	0.02%
c. Mobile	
6200-6250 kc	
8230-8330 kc	
11000-	
12340-12500 kc	0.05%
16460-16660 kc	
22000-22200 kc	
Other frequencies	0.02%
d. Aircraft	0.025%
e. Broadcasting	0.005%

This table is presented here so that the student will appreciate the necessity for accurate control and means of checking the operating frequency of a radio transmitter.

9-2. Amplitude-Modulated Radio Telephone Transmitters.- There are some design features that are nearly the same for all radio transmitters. Because of the high stability or low tolerance required of the various classes of radio transmitters, most transmitters employ quartz-crystal oscillators for their master frequency control. For many types of services the crystals, and in some cases the circuits, are kept at approximately constant temperature in a temperature-regulated box, but certain types of crystal cuts make temperature regulation unnecessary in some services. The oscillators are low power, generally not more than 7.5 watts, and are followed by buffer amplifiers operated as Class C, which increase the power level until the modulated stage is reached. Beyond the modulated stage the type of amplifier depends upon the class of the transmitter. Figure 9-1 illustrates a typical layout of a crystal oscillator and two buffer stages. Each buffer amplifier is operated as Class C and is tuned by a simple parallel circuit in the plate circuit. The manner in which one stage is coupled to the next varies and depends somewhat on whether the two stages are located near each other or some distance apart in the rack. Link coupling between stages is used to some extent when tuned circuits are employed in both the plate and grid circuits.

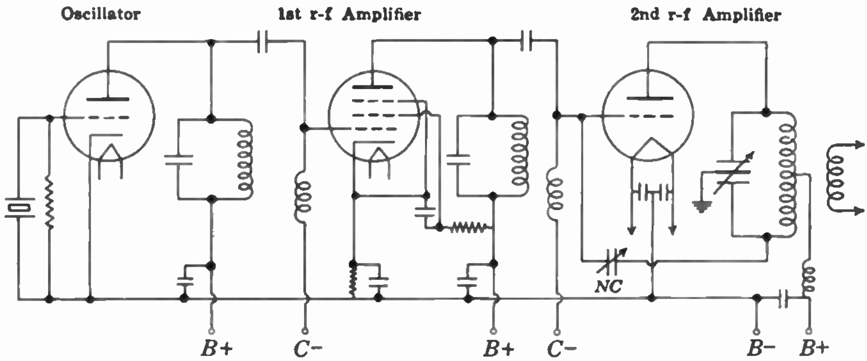


FIG. 9-1. A typical arrangement of the oscillator and buffer amplifier in an amplitude-modulated transmitter.

The buffer amplifiers in a transmitter may employ pentodes or neutralized triodes. The present-day practice seems to be to use pentodes because, if care is taken in laying out the circuits, they require no neutralization except at the high frequencies. When the operating frequency is higher than about 10 megacycles, the frequency is stabilized by a crystal oscillator operating at a sub-multiple of the operating frequency. Then one or more of the buffer amplifiers are operated as harmonic generators or frequency doublers or triplers.

All radio-telephone amplitude-modulated transmitters are essentially alike in several respects. There are, however, two different practices in regard to the location of the modulating amplifier with respect to the rest of the stages. One practice is to modulate the carrier at low power level and then to amplify the modulated wave with Class B or high-efficiency linear amplifiers to the desired power level for transmission. Figure 9-2 is a block diagram of low-level modulation. The other practice is to modulate at high power levels in the final stage or power amplifier. Consequently, high-level plate-circuit modulation requires a modulator that will furnish the full side-band power output of the transmitter. This means the audio-

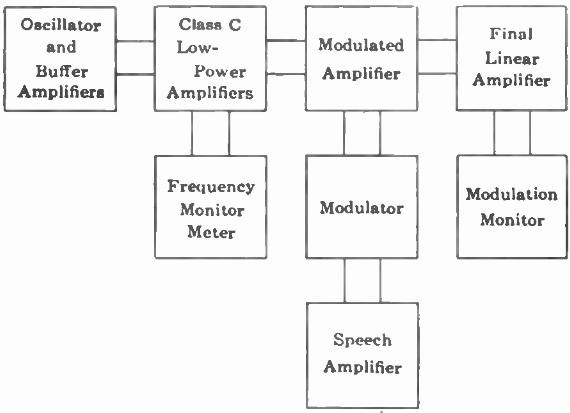


FIG. 9-2. Block diagram of a radio-telephone transmitter with low-level modulation.

power output of the modulator must be equal to one-half of the d-c power input of the modulated amplifier for 100 per cent modulation at an efficiency equal to that at which the carrier output is produced by the modulated amplifier; whereas for the low-level modulated type, the modulator must furnish side-band power that is only a fraction of the full side-band power output of the transmitter. In the low-level type the amplifiers, after modulation has been produced, must be linear, either Class B r-f amplifiers or a high-efficiency linear amplifier such as the Doherty.[1]

There are two practices for effecting modulation in transmitters; namely the plate-circuit-modulated Class C amplifier and the grid-circuit-modulated Class C amplifier. These two practices are illustrated in Figs. 9-3 and 9-4. Plate-circuit modulation requires more modulator power but the efficiency of power conversion by the modulated amplifier is higher and frequently the distortion may be made less. Also if the same type of tube were to be used in modulators of both types, the modulated power output would be very much greater in the plate-circuit modulator.

1. For the theory of the Doherty amplifier, see Doherty, "A New High Efficiency Power Amplifier for Modulated Waves," Proc. I.R.E., Sept. 1936; also Proc. I.R.E., Sept. 1939.

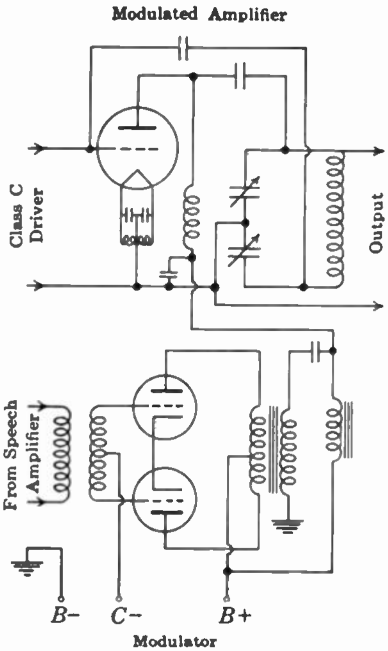


FIG. 9-3. Plate-circuit-modulated Class C amplifier.

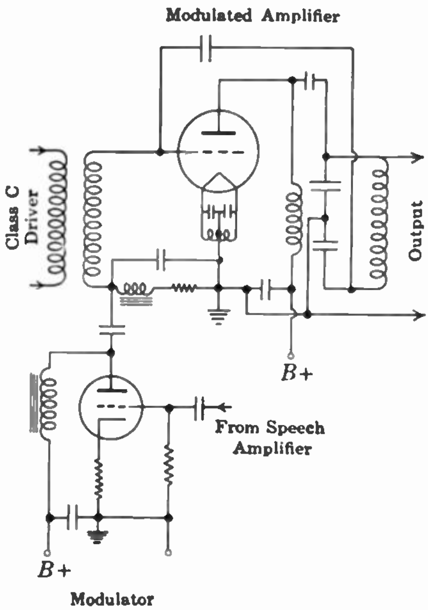


FIG. 9-4. Grid-circuit-modulated Class C amplifier.

The high-power stages of a radio-telephone transmitter generally comprise two tubes connected in push-pull. The modulator stage is often connected push-pull, since this type of operation gives a still better efficiency and a substantial reduction in even harmonics.

A radio-telephone transmitter for broadcast service must have a direct-reading frequency-deviation meter and a modulation monitor. The frequency-deviation meter reads in cycles off the assigned frequency, and is connected inductively to one of the stages ahead of the modulated amplifier. The modulation monitor reads the percentage of modulation and tells when the modulation goes over 100 per cent.

Modern radio-telephone transmitters employ inverse feed-back for reducing hum, noise, and distortion. One procedure is to impress a portion of the modulated output voltage on a diode detector and feed the rectified audio-frequency output of the diode into the input circuit of the speech amplifier or of the tube just ahead of the modulator. Feed-back is necessary to reduce hum when the filaments of the power tubes, particularly linear r-f amplifiers, are operated on a-c.

The higher power tubes of a radio transmitter are cooled either by forced air or water. In the water-cooled type, water is circulated around the anode of the tube by means of a close fitting water jacket. In the air-cooled type, fins are fitted to the anode and air may be forced up through these fins.

9-3. Frequency-Modulated Radio-Telephone Transmitters.- There are two or more types of frequency-modulated transmitters. The two most important types are the Armstrong and the reactance-tube modulator. It is beyond the scope of this chapter (see Ch. 6) to go into the details of modulating systems for these two ways of producing frequency modulation. However, in the Armstrong system, the carrier frequency is obtained directly from a crystal oscillator and therefore needs no indirect method for stabilization. A block diagram of the Armstrong system is shown in Fig. 9-5. The 200-kc frequency-modulated wave has a very low modulation factor m_f. The frequency multipliers step up both the frequency and the modulation factor.

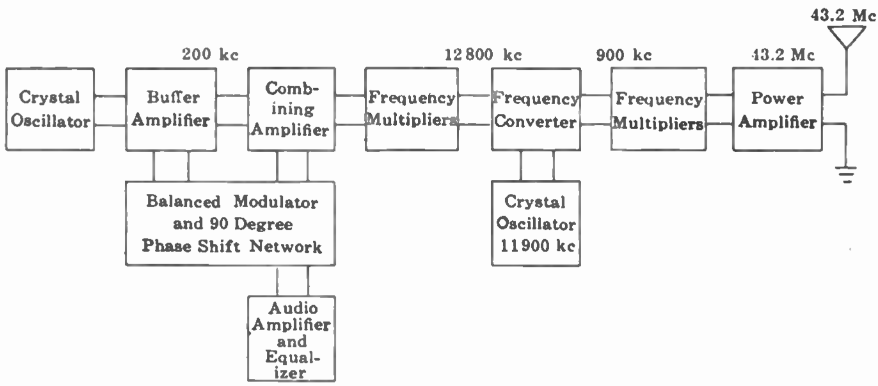


FIG. 9-5. Block diagram of Armstrong system for frequency modulation.

In the reactance-tube method, the oscillator is frequency modulated and cannot be directly crystal controlled. However, crystal control is obtained by a method quite similar to the frequency control in a radio receiver where a discriminator is used to control the normal adjustment of a variable-reactance tube. The block diagram is shown in Fig. 9-6. The frequency for normal performance of the discriminator shown is 1500 kc.

If the carrier frequency of the transmitter changes, because of a somewhat slow drift of the non-modulated or carrier frequency of the modulated oscillator, the input frequency to the discriminator is changed. This results in an output d-c voltage that shifts the bias of the variable-reactance tube in such a way as to bring the carrier frequency of the modulated oscillator back to its proper value. The variable-reactance tube

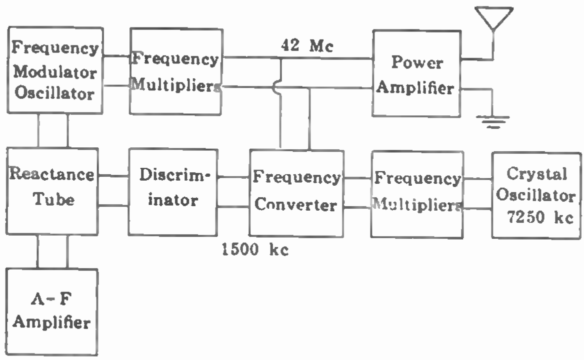


FIG. 9-6. Block diagram of a variable-reactance-tube method for frequency modulation.

serves two purposes. Its normal reactance is controlled by the discriminator, which keeps the carrier frequency of the modulated oscillator constant. The variation of its reactance is caused by the audio-frequency input, which in turn makes the reactance vary at an audio-frequency rate. The rate at which the reactance varies is proportional to the frequency of the audio signal, and the magnitude of variation is proportional to its amplitude.

In the Western Electric system[2] the carrier frequency of the modulated oscillator is held constant by a small motor mechanically tuning the condenser of the oscillator. This motor is energized by a frequency-divider system connected to the output of the modulated oscillator and a low-frequency crystal oscillator through a modulator. The output of this modulator produces a rotating magnetic field in the motor when the output of the frequency divider differs from that of the crystal-controlled oscillator. The direction of rotation depends upon which frequency is the higher. When the carrier frequency is correct the motor is at rest. The inertia of the motor prevents it from following the modulation frequency.

9-4. Code Transmitters.- A code transmitter does not differ materially from a radio-telephone transmitter except that frequently there are no modulated amplifier and modulator. All amplifiers are operated as Class C either at their fundamental frequencies or as frequency multipliers, depending upon the

2. Morrison, J.F.,"A New Broadcast-Transmitter Design for Frequency Modulation," Proc. I.R.E., Vol. 28, No. 10, October 1940.

frequency of operation. Keying is done by interrupting the excitation to one of the low-level amplifiers. This is accomplished by one of two methods. One method is to reduce the plate voltage of a low-level amplifier through a resistor that is also connected to a tube called the keying tube. The keying-tube plate current is controlled by changing the bias to its grid with the key. This changed bias greatly reduces the output of the tube and the excitation of the next tube. Hence, when the key is up, the r-f output of the transmitter is very low. When the key is down, the excitation is normal and the output is a maximum. The other method of keying changes the grid-bias voltage of a low-level tube. When the key is down, the bias voltage is normal and, when key is up, the bias voltage is large enough to reduce the r-f output almost to zero.

Code transmitters are apt to cause interference to other radio channels because of the high order side-band frequencies that are produced when the output changes rapidly from nearly zero to a maximum as the signal is made. The side-band frequencies can be suppressed by proper filter circuits between the output of the transmitter and the antenna, or prevented by the use of circuits in the keying system that do not let the energy level of the transmitter change so rapidly.

Short-wave code transmitters are generally crystal controlled. When the frequency is higher than about 10 megacycles, frequency-multiplier amplifiers are used in connection with a crystal-controlled oscillator that operates at a sub-multiple of the operating frequency. Long-wave code transmitters are commonly controlled by a master oscillator that is not crystal controlled.

Chapter 10

ULTRA-HIGH-FREQUENCY GENERATORS

One of the major problems encountered in the field of ultra-high-frequency techniques is the generation of ultra-high-frequency oscillations. This chapter will first discuss the factors that reduce the efficiency and power output and limit the frequency of operation of conventional vacuum-tube oscillators, and then treat special tubes and circuits that have been developed for the generation of waves with wavelength of the order of 20 cm or less in length.

10-1. Frequency Limit of Conventional Oscillators.- In order to increase the frequency of oscillation of conventional oscillators it is necessary to decrease the inductance and capacitance of the oscillating circuit. In the limit, the inductance is reduced to that of the grid and plate leads and the capacitance to that between the leads and the electrodes. The oscillating circuit is then physically reduced to a single loop, as shown in Fig. 10-1, composed principally of the grid and plate leads, which are bridged at their outer extremities by a condenser whose reactance is negligible at the frequency of oscillation. The oscillating circuit may be constructed as a Lecher wire system[1] having an equivalent length of a multiple of quarter wavelengths or of concentric lines whose equivalent lengths are a multiple quarter wavelengths. As the frequency is increased, it is found that a value

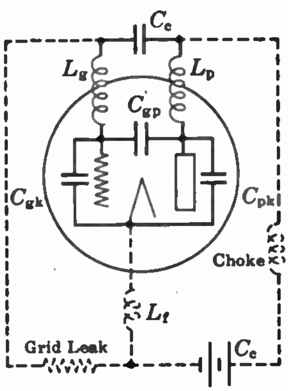


FIG. 10-1. Circuit of high-frequency negative-grid oscillator. The oscillatory circuit is indicated by solid lines. The reactance of C_c is negligible at the frequency of oscillation.

1. The term Lecher wire system used in this chapter refers to a two-wire transmission line, electrically long, but physically not large because of the short wavelengths at which it is used. Transmission lines are discussed in Ch. 11.

The material of this chapter is a logical sequence of the electron tube theory and circuits of the preceding chapters. Nevertheless, a thorough understanding of the content of this chapter will require some knowledge of the subject matter of Chs. 11, 12, and 14. Some readers will prefer to complete Chs. 11 to 14 before going through Ch. 10.

is reached, usually in the range between 10 and 60 megacycles, at which the efficiency and power output begin to fall off rapidly, and eventually oscillation ceases entirely. The loss of efficiency and power output, and the frequency limit are the result of one or more of the following factors:

A. Transit time of electrons between cathode and plate, which
   1. increases the effective grid conductance of the tube and
   2. shifts the phase of the plate current with respect to the grid voltage.

B. Limitation, by the physical structure of the tube, of the extent to which the parameters of the oscillating circuit can be reduced.

C. Increase of power loss in the oscillating circuit as the result of
   1. skin effect,
   2. large capacitance charging current, which results in large RI^2 loss,
   3. electromagnetic radiation from the circuit,
   4. dielectric losses in the tube base and envelope.

A. Electron transit time affects the efficiency and frequency limit of oscillation principally through its influence upon the input conductance of the tube. It has been shown, both theoretically and experimentally, that the input conductance (real part of input admittance), which is practically zero when transit time is unimportant, rises rapidly as the period of oscillation approaches the time of transit T_t of an electron between the cathode and the plate.[2] The input or grid conductance G_g of a negative-grid tube is given by the relation

$$G_g = Kg_m f^2 T_t^2 = Kg_m \frac{T_t^2}{T^2} \qquad (10\text{-}1)$$

in which K is a parameter depending upon the electrode spacing and upon the electrode voltages, g_m is the transconductance, f is the frequency, and $T = 1/f$ is the period of oscillation.[3]

2. W.R. Ferris, Proc. I.R.E., 24, 82, Jan. (1936), and accompanying paper by D.O. North.

3. The physical explanation for the real component of input admittance at high frequencies becomes apparent when it is noted that a portion of the input current is the charging current of the grid-plate capacitance. The current flowing into C_{gp} is caused to flow by the vector sum of the impressed grid-cathode voltage e_1 and the voltage e_0 developed across the plate load. At low frequency the load voltage across a resistive load is opposite in phase to the impressed voltage e_1 and adds algebraically to the impressed voltage in sending current through C_{gp}. Hence the resultant current is 90 degrees out of phase with the impressed voltage. When the

In tubes of conventional structure the transit time may be of the order of magnitude of 0.001 microsecond, which is the period of a 1000-megacycle wave.

Since the grid conductance either shunts the resonant tank circuit, or is coupled to it, increase of grid conductance increases the energy dissipation. The increased dissipation is objectionable not only because of the reduction of efficiency and power output, but also because the energy which is lost in the grid circuit appears at the plate in the form of heat and thus raises the temperature of the plate. The efficiency of operation is also affected by the shift in phase between the grid voltage and the plate current. This is because the circuit parameters must be changed in order to compensate for this phase shift.

Since the transit time decreases with an increase of plate voltage and with a decrease of plate-cathode spacing, the efficiency and the high-frequency limit of oscillation can be increased by raising the operating voltage or by using a tube with small electrode spacing.

B and C. Limitation of the upper frequency of oscillation by the minimum size to which the oscillating circuit can be reduced is a problem that may be solved to some extent by the use of tubes of special design. Reduction of lead length not only reduces circuit inductance and capacitance, but also circuit resistance. Losses caused by skin effect can be minimized by the use of conductors of low resistivity metals and large surface area. Conductor losses resulting from high charging currents necessitate the use of electrode structures that minimize interelectrode capacitance. Losses resulting from radiation can be minimized by close spacing of leads (of the order of 1/100 of the wavelength), but decrease of conductor spacing below a certain value increases the r-f resistance of the conductors. A compromise must therefore be made in choosing the lead spacing. Dielectric losses in the tube envelope are objectionable not only because of the reduction of output and efficiency, but also because they may cause disintegration of tube seals. Dielectric losses are reduced by eliminating the tube base and making connections directly to the leads and by bringing the conductors through the glass at points on the conductors where there are potential nodes.

The low interelectrode capacitances requisite for high

Footnote Continued

period of the impressed voltage approaches the transit time, however, the plate current, and hence the voltage across a resistive load, begins to lag the impressed voltage. The resultant voltage across C_{gp} therefore lags the impressed voltage and the charging current has a component in phase with e_1.

resonance frequency and for low charging-current loss must be attained by the use of small, carefully designed electrodes. Although interelectrode capacitances can also be reduced by increasing the electrode spacing, this has the objectionable effect of increasing electron transit time. It has been shown that for optimum conditions of operation, the dimensions of the tube must be decreased in proportion to the operating wavelength. If all the linear dimensions of a vacuum tube are divided by a factor n, the tube factors and the plate current at fixed operating voltages remain unchanged, but the interelectrode capacitance lead inductance, and transit time are divided by the factor n.[4] From this fact it follows that if a good design at one wavelength is available, the optimum procedure in redesigning at a lower wavelength is to reduce all tube and circuit linear dimensions in proportion to the wavelength. The allowable plate dissipation and available emission, however, are inversely proportional to n^2 and current densities directly proportional to n^2. The power output that can be obtained from tubes of small dimensions is therefore less than that obtainable from larger tubes.

Dependence in a given tube of limiting frequency of oscillation upon the plate voltage may be taken as an indication that this limit is the result of electron transit time, rather than of the physical size of the circuit or of circuit losses.

10-2. Ultra-high-frequency Negative-grid Tubes.- The low transit time, small lead inductance and capacitance, and small interelectrode capacitance requisite to the production of ultra-high-frequency oscillations have been attained in the "acorn"[5] and "door-knob"[6] types of tubes. Acorn type triodes will oscillate at wavelengths as low as 40 cm. Because acorn tubes are designed principally for use in amplifier and low-power oscillator circuits, they will not be discussed further.

Figure 10-2 shows the Western Electric type 316A door-knob tube, which has an upper frequency limit of about 700 megacycles. Unusual features of the design of this tube include the complete elimination of the usual "press" and the usual close spacing of leads, the shortness of the leads, and the small size of the electrodes. The grid is in the form of a number of straight wires parallel and equidistant from the axial filament and supported by cooling collars at each end. The upper frequency limit of this tube is the result of electron transit time.

Figure 10-3 shows the structure of the Western Electric

---

4. Salzberg, B. and D.G. Burnside, Proc. I.R.E., 23, 1142 Oct. (1935).
5. Salzberg, B., Electronics, Sept. (1934); Salzberg, B., and Burnside, D.G., Proc. I.R.E., 23, 1142 Oct. (1935).
6. Samuel, A.L., Jour. App. Physics, 8, 677 Oct. (1937).

FIG. 10-2. Western Electric type 316A tube.

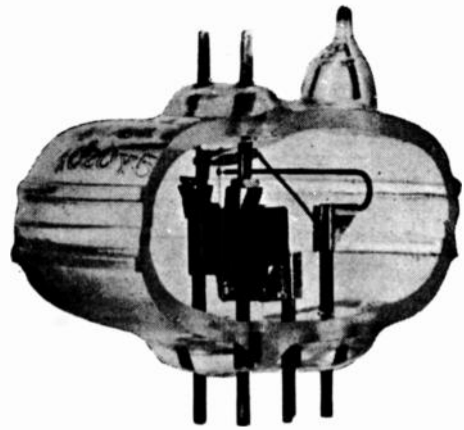


FIG. 10-3. Western Electric Type 368A Tube.

type 368A tube, which has a high-frequency limit of approximately 1700 megacycles. In this tube the grid and plate elements are supported by wires that go straight through the tube envelope and thus provide two independent paths to each of these elements. The tube has only one set of filament leads, one of which is very short. The grid consists of a series of tungsten wire loops supported by a cooling fin and projects into a slot in the block of graphite that acts as the plate element. The limiting factor in the power dissipating ability of this tube is the temperature of the grid, which is unusually close to the filament and may emit electrons if it becomes too hot.

The double-ended construction of the 368A tube makes possible the use of this tube in an oscillating circuit in which the tube elements are at the center of a half-wave Lecher system with closed ends. Because the lumped capacitance between the grid and plate may be assumed to be divided between the two quarter-wave halves of the Lecher system, the frequency of oscillation is higher than in a single-ended quarter-wave system, in which the total capacitance is associated with the single quarter-wave circuit. The double-ended arrangement decreases radiation losses and, because only half the charging current flows through each set of leads, it decreases the resistance loss in the leads.

10-3. Ultra-high-frequency Negative-grid Oscillators.- Figure 10-4 shows the circuit of an oscillator using the 368A tube. This circuit is tuned by means of concentric line "stubs" (see Ch. 11), which have the characteristic advantages over Lecher wire systems and other open circuits possessed by concentric lines. Because the electromagnetic fields are confined to the inside of the tubing, radiation from these portions of the circuit is prevented. Furthermore, when they are tuned so that there is a current node at the open end, the high-frequency current is confined to the inside of the lines. This simplifies the problem of applying direct voltages to the elements without

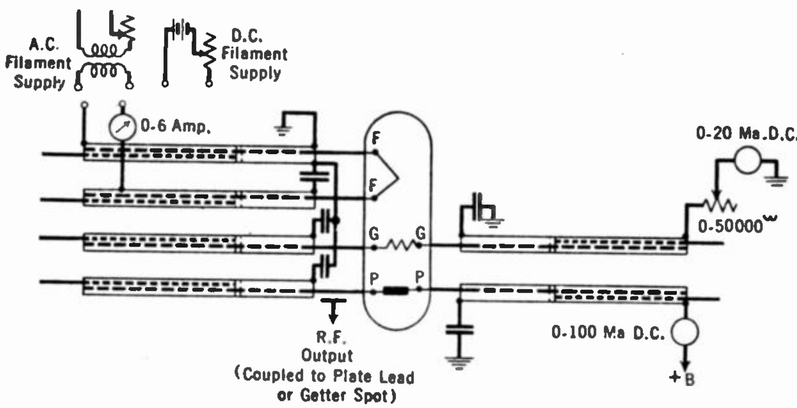


FIG 10-4. Circuit of Ultra-High Frequency Oscillators.

allowing high-frequency currents to flow in the supply lines. A third advantage is the ease with which the circuit can be tuned. Since the fields are confined to the inside of the stubs, body capacitance has no effect upon the tuning.

Because of the coupling between the filament and the grid and plate, tuning stubs must be used not only in the grid and plate leads but also in the filament leads in order to prevent the flow of high-frequency currents in the filament supply lines.

Although the highest frequency at which the 368A tube will oscillate is too low for many ultra-high-frequency applications, appreciable output at wavelengths as low as 9 or 10 cm can be obtained by extracting power at the second harmonic of the fundamental frequency of oscillation. In order to filter out the fundamental output, the vacuum tube is mounted, as shown in Fig. 10-5, within a brass tube whose diameter is sufficiently large to allow it to act as a wave guide for the propagation of second-harmonic power but not large enough to allow the propagation of fundamental power (see Ch. 14). For the production of 9- or 10-cm waves a tube having an inside diameter

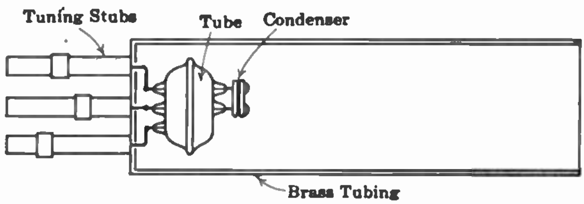


FIG. 10-5. 10-cm negative-grid harmonic oscillator.

of 2 1/2 or 3 inches is satisfactory. The energy may be radiated from the open end of the brass tube, conducted along a wave guide extension of the tube or it may be taken from the brass tube at a distance of 10 inches or more from the vacuum tube by coupling electrostatically or electromagnetically to concentric cables (see Ch. 14). Figure 10-6 shows the circuit of the oscillator. The condenser C, the reactance of which is negligible at 1500 Mc, is connected between the ends of one set of grid and plate leads which, with the grid and plate, form a quarter-wave Lecher system closed at the condenser end. The other set of grid and plate leads is adjusted with a concentric line stub to an effective length of one-fourth or three-fourths the wavelength. Radiation takes place from portions of the oscillating circuit other than the tuning stub. To prevent loss of energy through the filament supply lines, concentric line stubs are also used in the filament leads. The by-pass condensers shown in Fig. 10-6 consist of four brass disks, separated by mica, as shown in Fig. 10-7. These disks, which are located at the vacuum-tube end of the brass tube, support the concentric line stubs, the center conductors of which are attached to the tube leads. The filament, grid, and plate supply lines, indicated by the letters F, F, G, and P in Fig. 10-6, are also connected to the disks by means of soldering lugs or other convenient connectors.

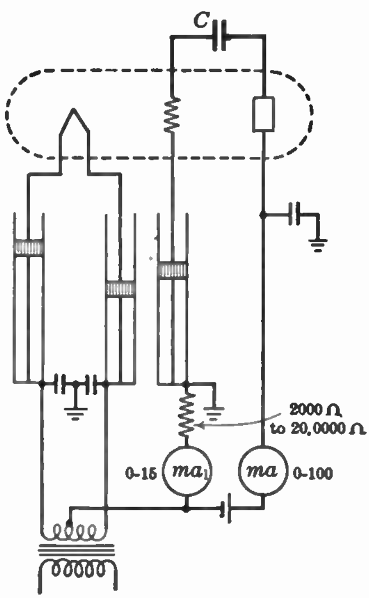


FIG. 10-6. Circuit diagram of 10-cm harmonic oscillator.

A 9-cm output of sufficient amplitude for most ultra-high-frequency experiments is obtained with a plate voltage of 275 to 300 volts, and a grid resistor of about 10,000 ohms. The output can be increased by decreasing the grid resistor to lower values and by raising the plate voltage, but the plate dissipation must be kept within the allowable value of 20 watts. As the adjustment of the tuning stubs is likely to be very critical, it is advisable to use reduced plate voltage and a protective resistance of 500 to 1000 ohms in the plate circuit until the circuit oscillates.[7] A good indication of oscillation is the flow of grid current, although a detector suitable for 9-cm waves should also be used. There are usually several combinations of settings of the tuning stubs at which the circuit will oscillate, but one setting gives the highest second-harmonic output. When the oscillator is properly tuned, very little high-frequency radiation can be picked up from the supply lines. The output is affected considerably by changes of filament current.

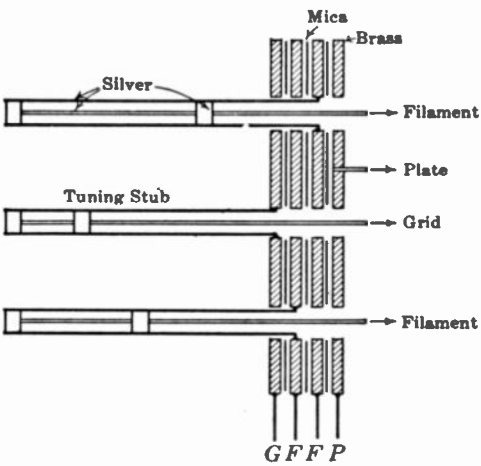


FIG. 10-7. Structure of tuning stubs and by-pass condensers for 10-cm harmonic negative-grid oscillator.

10-4. Positive-grid Oscillators.- High-frequency oscillations dependent upon electron transit time may be produced by applying positive voltage to the grid of a high vacuum triode and zero or a small negative voltage to the plate. As in ultra-high-frequency negative-grid oscillators, the principal oscillating circuit usually consists of the grid and plate leads, which form a Lecher wire system. The principal oscillating circuit may, however, be placed either between the grid and the cathode or the plate and the cathode. The frequency of oscillation is usually dependent upon the values of the parameters of the oscillating circuits, but under certain conditions the frequency of

7. The writer uses a 500-ohm resistor in the plate supply line in adjusting the circuit and reduces the plate voltage to a value at which the plate current does not exceed 50 ma when the tube is not oscillating (about 200 volts).

oscillation over a limited range of circuit parameters is independent of the external circuit and depends only upon the transit time of the electrons from the cathode to the plate. Oscillations independent of circuit parameters were first observed by Barkhausen and Kurz[8] in 1920, and bear the names of these discoverers. Oscillations dependent upon circuit parameters are called Gill-Morrell oscillations, after the men who first observed them.[9] The frequency of Gill-Morrell oscillations cannot be varied continuously by varying only the circuit parameters or only the electrode voltages, since the circuit oscillates only when the transit time is properly related to the period or periods of oscillation of the circuit.

10-5. **Theory of Positive-grid Oscillators.**- The operation of a positive-grid oscillator may be explained qualitatively by studying the motion of individual electrons between the cathode and the plate (A) when the grid and plate potentials are constant, and (B) when one or both have an alternating component relative to the cathode whose period is equal to the transit time of an electron from the cathode to its point of closest approach to the plate.

A. Assume first that the grid and plate potentials are constant and that electrons leave the cathode with zero initial velocity. An electron at the cathode is accelerated by the field between the cathode and the grid. When it reaches the plane of the grid it may strike the grid. It may, however, pass between the grid wires into the grid-plate space, where it will be slowed down by the grid-plate field. If the plate is slightly negative relative to the cathode, the electron will come to rest (in the zero potential plane) before it strikes the plate and will then again be accelerated toward the grid. If it again passes between the grid wires, it will be slowed down by the grid-cathode field and will come to rest at the cathode. If it does not enter the cathode it may again move toward the grid. Fig. 10-8 shows the approximate manner in which the position of the electron varies with time.[10]

B. Assume now that there is superimposed upon the steady grid voltage a sinusoidally varying voltage having a period equal to one-half the time taken for an electron to move from the cathode to the vicinity of the plate and back to the cathode. Consider first an electron that leaves the cathode at the instant when the alternating component of grid voltage is zero and the

8. Barkhausen, H., and Kurz, K., Phys. Zeits., 21, 1 (1920).
9. Gill, E.W.B., and Morrell, Phil. Mag., 44, 161 (1922).
10. Oscillators of any type in which electrons swing back and forth, as illustrated in Figs. 10-8 and 10, are sometimes called "pendulum-type oscillators."

grid potential is changing from negative to positive. The approximate manner in which the position of the electron varies with time is shown in Fig. 10-9. During the time that the electron moves from the cathode to the grid, the grid is more posi-

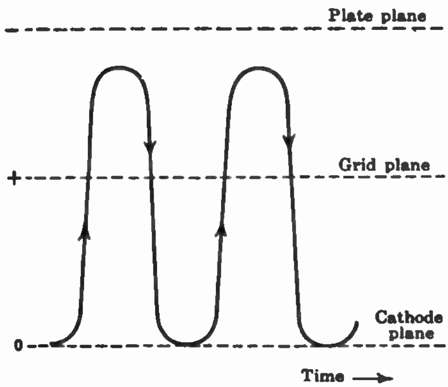


FIG. 10-8. Motion of an electron in a positive-grid tube with constant electrode voltages.

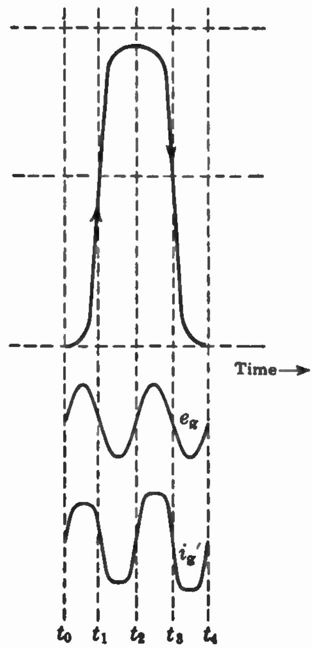


FIG. 10-9. Motion of an electron that absorbs energy from source of alternating grid voltage.

tive than it would be without the alternating component of grid voltage, and so the acceleration at every instant in this portion of the path is greater than it would be with constant grid voltage. At the instant the electron reaches the plane of the grid the alternating component of grid voltage reverses, and so the retarding field between the grid and the plate during the interval $t_1 - t_2$ is lower than with steady grid voltage, and the retardation of the electron at every instant in this interval is lower. Because of the greater acceleration between the cathode and the grid and the lower deceleration between the grid and the plate, the electron approaches closer to the plate than it does with steady grid voltage, and may actually strike the plate. If it does not strike the plate, it returns to the grid or cathode. The difference in potential between the grid and plate during the interval $t_2 - t_3$, however, is greater than that during the interval $t_1 - t_2$, and the difference in potential between the grid and the cathode is less during the interval $t_3 - t_4$ than during the interval $t_0 - t_1$. Consequently the acceleration throughout the interval $t_2 - t_3$ exceeds the deceleration throughout the interval $t_1 - t_2$, and the deceleration throughout the interval $t_3 - t_4$ is less than the acceleration throughout the interval $t_0 - t_1$. The electron therefore reaches the cathode with appreciable velocity, and will of necessity enter the cathode and be unable to make a second trip to the plate. The excess energy with which the electron strikes the cathode must come from the source of alternating grid voltage.

Figure 10-10 shows the manner in which the position of an electron varies with time when it leaves the cathode at the instant when the alternating grid voltage is zero and changing from positive to negative. In this case the acceleration between the cathode and the grid is lower than with constant grid voltage and the deceleration between the grid and the plate higher, and so the electron does not approach the plate so closely as with constant grid voltage. Furthermore, the acceleration throughout the interval $t_2 - t_3$ is less than the deceleration during the interval $t_1 - t_2$, and the deceleration throughout the interval $t_3 - t_4$ exceeds the acceleration throughout the interval $t_0 - t_1$. Consequently the electron comes to rest before it reaches the cathode. The fact that the electron comes to rest before it reaches the cathode, but would still have considerable energy at the same distance from the cathode if the grid voltage were constant, indicates that the electron has given up energy to the source of alternating grid voltage.

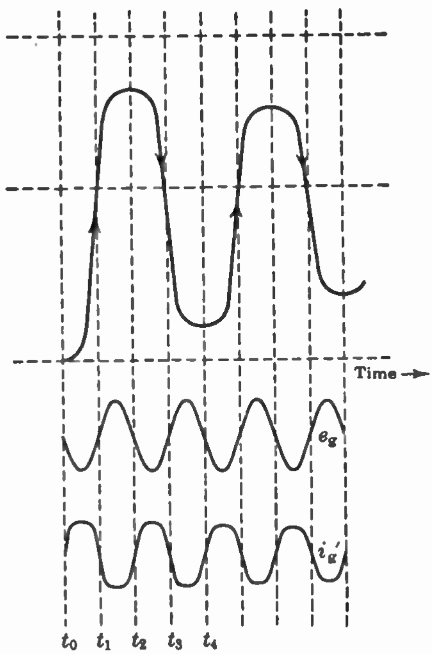


FIG. 10-10. Motion of an electron that gives up energy to the source of alternating grid voltage.

After the electron comes to rest it may make another trip toward the anode. This time, however, it starts from a point the potential of which is positive relative to the cathode and will consequently gain less energy in moving to the plane of the grid than when it moved from the cathode to the grid during the first trip. It therefore comes to rest at a point farther from the plate than during the first trip and will return to a point farther from the cathode than at the end of the first trip. Although it is possible that the electron may continue to oscillate about the grid with decreasing amplitude, each time it passes through the grid it is drawn closer to one of the grid wires by an amount that increases during each succeeding transit because of the reduction of the velocity with which it passes through the plane of the grid. For this reason it is improbable that the electron will make a large number of excursions before striking the grid. A large number of oscillations is, in fact, undesirable, since the transit time decreases as the amplitude of oscillation decreases.[11] For this reason there is an advance in the phase of the oscillation of an electron relative to that of the alternating voltage. As the difference in phase increases, the energy delivered by the electron during an oscillation becomes smaller and eventually becomes zero. In order to make the net energy delivered to the source of alternating voltage as large as possible, the electron must be removed from the space before the phase shift becomes so great that the electron takes energy from the source.

Electrons may leave the cathode at any instant in the cycle of alternating voltage. Analyses similar to those made above show that any electron that leaves the cathode at any instant in the cycle when the alternating grid potential is

11. Let s be the amplitude of oscillation of an electron, i.e., the maximum distance it moves toward the cathode from the grid. If the cathode-grid field is assumed to be uniform, the acceleration a of an electron in the cathode-grid space is

$$a = \frac{e}{md}\left[E_g + E_m \sin(\omega t + \theta)\right] \qquad (a)$$

in which E_g is the direct grid voltage relative to the cathode, E_m is the amplitude of the alternating grid voltage, e and m are the charge and mass of an electron, and d is the cathode-grid spacing. The velocity of an electron at a time t after it starts moving toward the grid from its point of greatest distance from the grid is

$$v = \int_0^t a\,dt = \frac{e}{md}\left[E_g t - \frac{E_m}{\omega}\cos(\omega t + \theta) + \frac{E_m}{\omega}\cos\theta\right] \qquad (b)$$

The distance moved in the time t is

$$x = \int_0^t v\,dt = \frac{e}{md}\left[\frac{1}{2}E_g t^2 - \frac{E_m}{\omega^2}\sin(\omega t+\theta) + \frac{E_m}{\omega^2}\sin\theta + \frac{E_m t}{\omega}\cos\theta\right] \qquad (c)$$

Although eq. (c) cannot be directly solved for t, the form of the equation indicates that the time taken by an electron to move through the distance s to the plane of the grid from its point of greatest departure from the grid decreases with decrease of s. In a similar manner it may be shown that the time taken for the electron to move from the grid to its point of closest approach to the plate decreases with the amplitude of oscillation of the electron.

changing in the positive direction gains energy from the source of alternating grid voltage and will either strike the plate after its first crossing or the cathode upon its first return. All electrons that leave at any instant when the alternating grid voltage is changing in the negative direction, on the other hand, deliver energy to the source of alternating voltage and may make a number of excursions, during each of which they deliver energy to this source. Since the energy gained by an electron of the first group may be shown to equal approximately the energy lost during the first excursion of an electron of the second group that leaves one-half cycle later,[12] and since the number of electrons that leave the cathode in any interval of given length may on the average be assumed to be constant, it follows that more energy is delivered to the source of alternating grid voltage than is taken from it. The source of alternating grid voltage therefore gains energy, which must be supplied by the source of direct grid voltage.

The student may show that an alternating plate voltage, opposite in phase to the alternating grid voltage, acts upon a given electron in a manner similar to the alternating grid voltage. When an oscillatory circuit is connected between the grid and the plate, alternating components of grid and plate voltage, opposite in phase, are supplied by the circuit. If the energy gained by the oscillatory circuit as the result of the motion of the electrons in the tube exceeds the energy dissipated in the

12. If the amplitude of the alternating grid voltage is small in comparison with the direct grid voltage, the time of the first transit of an electron is nearly equal to the transit time for zero alternating grid voltage. Under this assumption, eq. (b) of the preceding footnote shows that the velocity of an electron at the plane of the grid during its first excursion is

$$v_g \doteq \frac{e}{md}\left[E_g t_o - \frac{E_m}{\omega}\cos(\omega t_o + \theta) + \frac{E_m}{\omega}\cos\theta\right] \qquad \text{(d)}$$

in which t_o is the time taken by an electron to move to the grid from its point of greatest distance from the grid. The second and third terms of eq. (d) represent the velocity gained by the electron as the result of the alternating grid voltage. Since changing θ by 180 degrees is equivalent to reversing the algebraic signs of these two terms, it follows that the magnitude of the change in velocity of an electron as the result of the alternating voltage is, at least to the first order of approximation, the same for two electrons whose time of departure from the cathode differs by half a cycle. Hence the magnitude of the energy delivered to or taken from the source of alternating voltage is equal for two such electrons. A similar analysis shows that the magnitude of the energy transfer between electrons and the source of alternating voltage during the first transits of the electrons from the grid toward the plate is equal for electrons whose time of departure from the grid differs by a half a cycle.

cricuit, sustained oscillation will take place. Sustained oscillations may be produced by a similar process if the oscillatory circuit is connected between either the grid or the plate and the cathode.

10-6. The Positive-grid Oscillator Tube Considered as a Negative Resistance.- The fact that sustained oscillation takes place when a resonant circuit is connected between two of the electrodes suggests that between these points the tube acts like a negative resistance. This can be shown to be so. As electrons approach the surface of an electrode they induce a positive charge on the surface, i.e., they repel electrons from the surface within the metal. Similarly, as space electrons move away from an electrode, electrons within the electrode flow toward the surface of the electrode. As a result, electrons move in the external circuit and thus produce a current (not sinusoidal),[13] the conventional direction of which is opposite to the direction of motion of the electrons. (Conventional direction of current is the direction in which positive charges would move.) This current is superimposed upon the current resulting from the flow of electrons that actually strike and enter the electrode and upon the current that flows as the result of interelectrode capacitances. The lower curves in Figs. 10-9 and 10-10, derived from the upper curves, show the general form of the grid current flowing in the external circuit as the result of the motion of the two electrons to which these figures apply. Comparison of the waves of plate voltage and plate current shows that the fundamental component of the induced grid current resulting from the motion of the electron that takes energy from the source of alternating grid voltage is in phase with the alternating grid voltage, whereas that resulting from the motion of the electron that delivers energy to the voltage source is opposite in phase to the grid voltage. Similar analyses indicate that all electrons that leave the cathode at such instants that they take energy from the source of alternating grid voltage produce fundamental induced grid currents having components in phase with the grid voltage, whereas all electrons that leave at such instants that they deliver energy to the source of alternating voltage produce fundamental induced grid currents having components opposite in phase to the grid voltage. Since electrons of the former type leave the interelectrode space after not more than one excursion, whereas electrons of the latter type may make several excursions, the total number of the second type that are in the space exceeds the number of the first type, and so the resulting 180 degree out-of-phase fundamental grid current exceeds the resulting in-phase fundamental grid current. The net fundamental grid current resulting from the oscillation

13. Shockley, W., Jour. App. Physics, 9, 635 Oct. (1938).

of electrons is therefore opposite in phase to the grid voltage, and the tube acts like a negative resistance. A similar analysis shows that the tube also acts like a negative resistance between the plate and the cathode or between the grid and the plate.

10-7. Frequency of Oscillation of Positive-grid Oscillators.- A negative tube resistance may also be shown to exist when the period of the alternating electrode voltages differs slightly from the cathode-plate transit time. As the frequency difference is increased, however, the magnitude of the negative resistance increases. Since, as shown in Sec. 4-17, the magnitude of the negative resistance must be less than the effective resistance of the oscillating circuit in order for sustained oscillation to be possible, the circuit oscillates only over a limited range of frequency in the vicinity of the transit frequency. As the difference between the transit frequency and the frequency of the resonant circuit is increased, the amplitude of oscillation decreases and finally becomes zero. It may be shown, furthermore, that there are other frequency ranges in which the tube acts like a negative resistance and sustained oscillation may occur. Because the magnitude of the negative resistance in these ranges is greater and the energy dissipation of the circuit may be higher, these other ranges of oscillation are usually not observed. "Dwarf" oscillations, of frequency higher than the transit frequency have, however, been reported by a number of investigators.[14]

A Lecher wire system connecting two of the electrodes can resonate to the transit frequency when its effective length is equal to $(n/2 - 1/4)\lambda_t$, in which n is any integer and λ_t is the wavelength corresponding to the transit frequency. (See Table 11-II of Ch. 11.) As the length is increased, therefore, it passes through successive ranges in which sustained oscillations are obtained. Increased loss in the circuit, however, causes the amplitude of oscillation and the range in which oscillation takes place to be less in each succeeding range. Since the grid-plate, grid-cathode, and plate-cathode circuits are coupled together through the interelectrode capacitances, the oscillating system is in general complicated and may have a number of natural frequencies of oscillation. It is possible to find combinations of circuit parameters such that over limited ranges the frequency of oscillation is not affected by changes of parameters. The Barkhausen-Kurz type of oscillation is then observed. Multiple oscillation frequencies can be avoided by the use of concentric line stubs in all electrode leads, as in the negative-grid oscillator circuit of Fig. 10-4.

14. See for instance, W.D. Hershberger, Proc. I.R.E., 24, 964 July (1936).

10-8. Difficulties in the Design and Operation of Positive-grid Oscillators.- Because most of the electrons leaving the cathode are caught by the grid before they leave the cathode space, the ratio of the alternating grid current to the direct grid current is small. Hence the efficiency of the positive-grid oscillator is low. This is objectionable not only because of the resulting low power output, but also because of the difficulty of dissipating the energy released at the grid. The problem is particularly difficult at wavelengths of the order of 20 cm and less, since short transit time and low inter-electrode capacitance necessitate the use of high grid voltage, close spacing between electrodes, and small electrode area. For this reason positive-grid oscillators have not as yet proved to be of great value in this frequency range. The theory of operation of these oscillators is, however, of importance since it may suggest modifications or new types of tubes that operate more efficiently.

10-9. Magnetron Oscillators.- One of the most satisfactory generators of ultra-high-frequency oscillations is the magnetron, which consists, in its most common form, of a two-element tube with cylindrical plate structure and a coaxial filament, used with a magnetic field parallel, or nearly parallel to the filament. Although an alternating magnetic field may be used in the production of oscillations of comparatively low frequency, the magnetic field is maintained constant in the generation of ultra-high-frequency oscillations. As first described by Hull,[15] the anode consisted of a continuous cylinder. The anode in the type of magnetron now commonly used is divided into two independent semi-cylindrical segments, as shown in Fig. 10-11, which are connected through an oscillatory circuit which may be either entirely inside of the tube envelope, or partially outside.

Two types of oscillations are obtained with magnetrons using constant magnetic field. The first type is dependent upon the fact that the static current-voltage characteristic of either anode of a split-anode magnetron has a portion with negative slope. A magnetron that generates this type of oscillation is usually called the "negative-resistance" or "dynatron" magnetron. The negative-resistance magnetron operates efficiently only at frequencies that are low in comparison with the transit frequency. The frequency of oscillation can be varied continuously by changing the constants of the oscillatory circuit. The second type of oscillation, which is in some respects similar to that obtained in positive-grid oscillators, depends upon the transit time of the electron from the cathode to the anode.

15. Hull, A.W., Phys. Rev., 18, 31 July (1921).

Oscillation of this type takes place only when the magnetic and electric fields are adjusted so that the time taken for the electron to move from the cathode toward the anode and back is equal to, or nearly equal to, the period of oscillation of the oscillatory circuit. Magnetrons generating this type of oscillation are called "transit-time magnetrons."

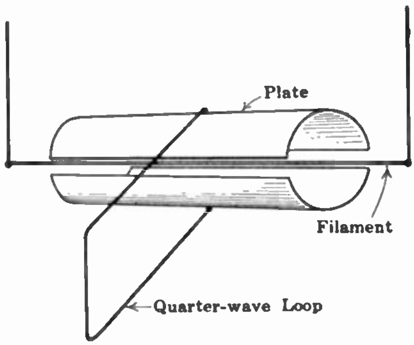


FIG. 10-11. Electrode structure of split-plate magnetron.

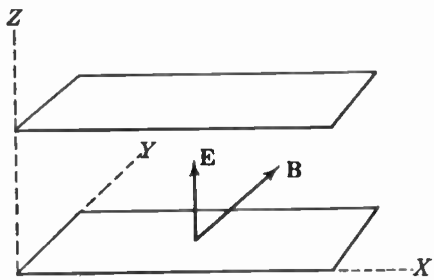


FIG. 10-12. Electric and magnetic fields in parallel-plate magnetron.

10-10. Paths of Electrons in Magnetrons under Static Conditions.- Although practical magnetrons do not make use of plane parallel electrodes, it is instructive to investigate the motion of electrons in theoretical magnetrons of this type. Fig. 10-12 shows two plane parallel electrodes which will be assumed to be of infinite extent and to lie parallel to the *X*-*Y* plane of coordinates. The magnetic flux is assumed to be uniform and to be parallel to the *Y*-axis. The electric field is assumed to be uniform and to be parallel to the *Z*-axis. The equations of motion of an electron whose instantaneous position in the space between the electrodes is specified by the coordinates x, y, z are

$$m \frac{d^2 z}{dt^2} = Ee - \frac{Be}{c} \frac{dx}{dt} \tag{10-2}$$

$$m \frac{d^2 x}{dt^2} = \frac{Be}{c} \frac{dz}{dt} \tag{10-3}$$

$$\frac{d^2 y}{dt^2} = 0 \tag{10-4}$$

in which m is the mass and e the magnitude of the charge of an electron in esu, E is the electric field strength in esu, B is the flux density in gauss, and c is the velocity of light in cm/sec. Solution of eqs. (10-2) and (10-3) yields the following equations for the path of an electron initially at rest at the origin:

$$x = \frac{cE}{B\omega}(\omega t - \sin \omega t) \tag{10-5}$$

$$z = \frac{cE}{B\omega}(1 - \cos \omega t) \tag{10-6}$$

in which

$$\omega = Be/mc \tag{10-7}$$

Equations (10-5) and (10-6) are those of a cycloid generated by a point on a circle of radius $E/B\omega$ rolling on the plane of the cathode with an angular velocity ω. The maximum distance to which the electron moves normal to the cathode is $2cEm/B^2e$. When this distance is just equal to the anode-cathode spacing, the electrons just graze the surface of the plate and the plate current is just cut off. For an electrode spacing d, the critical relation between the magnetic flux density and the plate voltage for plate-current cutoff is

$$B^2 = 11.32\ E_b/d^2 \tag{10-8}$$

in which B is measured in gauss, E_b in volts,[16] and d in centimeters. When B is less than the value given by eq. (10-8), the electrons strike the plate. When B exceeds this value, on the other hand, they return to the cathode and may make other excursions toward the plate. Figure 10-13 shows possible types of

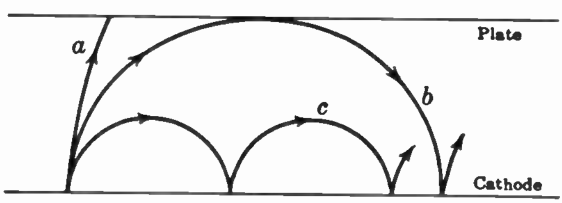


FIG. 10-13. Paths of electrons in parallel-plate magnetron with static fields.

electron paths. It should be noted that eq. (10-7) indicates that the angular velocity of the electrons, and hence their time of transit, depends only upon the strength of the magnetic field.

The analysis for a magnetron with concentric cylindrical electrodes is extremely complicated in comparison with that for the magnetron with plane parallel electrodes. A complete analysis for the single-anode cylindrical magnetron has been made by Brillouin.[17] This analysis shows that the angular velocity of electrons leaving the filament increases gradually up to a limiting value

16. The reader should distinguish between E for electric field intensity and E for voltage.
17. Brillouin, L., Phys. Rev. 60, 385 Sept. 1, (1941).

$$\omega_m = 8.84 \times 10^6 B_{gauss} \qquad (10\text{-}9)$$

The electrons describe spiral paths in a plane normal to the filament after they reach a distance L from the filament given by the relation

$$L^2 = 2.3 \times 10^3 I_b/hB^3 \quad \text{cm} \qquad (10\text{-}10)$$

in which I_b is the anode current in milliamperes, h is the length of the filament in cm, and B is the flux density in gauss. The anode current is just reduced to zero when the following relation holds between the flux density and the anode voltage:

$$B = 6.7\ E_b^{1/2}\ b/(b^2 - a^2) \qquad (10\text{-}11)$$

in which b is the anode diameter in cm and a the filament diameter. The electrons then rotate about the filament with an angular velocity given by eq. (10-9). In practical diodes the filament diameter is so much smaller than the plate diameter that eq. (10-11) reduces to[15]

$$B = 6.7\ E_b^{1/2}/b \qquad (10\text{-}12)$$

When the flux density is just above the cutoff value, an electron makes approximately 0.7 of a revolution about the filament in moving from the cathode to the vicinity of the plate and back. Figure 10-14 shows typical electron paths.

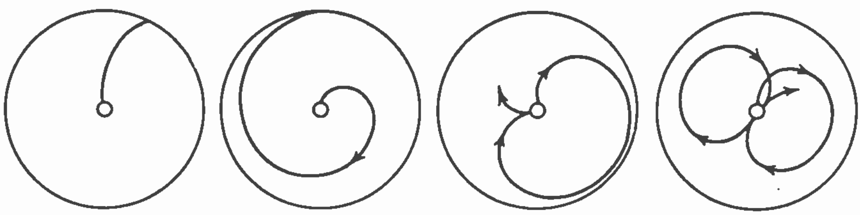


FIG. 10-14. Electron paths in a cylindrical magnetron with static fields.

The above analysis neglects an important effect that Linder has shown to exist, namely, the conversion of some of the orbital energy of the electrons into energy of random motion.[18] The most probable cause of this phenomenon appears to be collisions between electrons moving away from and returning toward the cathode. It results in the formation of a region about the cathode in which electron current is limited by space charge. The boundary of this region acts as a virtual cathode. Because of their energy of random motion, some electrons may

18. Linder, E.G., Proc. I.R.E., 26, 346 March (1938).

return to the actual cathode with considerable excess energy. The resulting bombardment may produce an objectionable rise in cathode temperature. This effect will be discussed in Sec. 10-14. A second consequence of the energy of random motion is the flow of some plate current when the flux density exceeds the theoretical cutoff value indicated by eqs. (10-8) and (10-12).

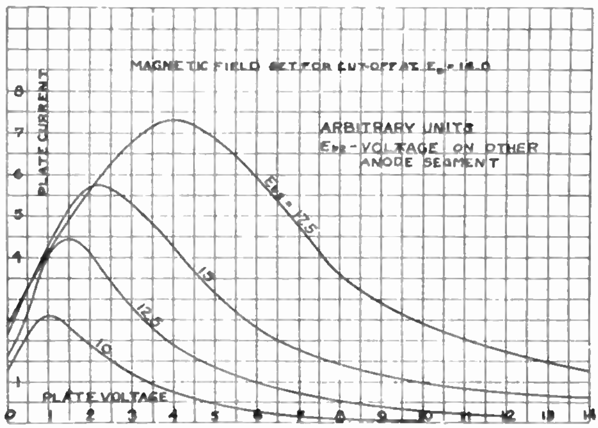


FIG. 10-15. Current vs voltage for one segment of a split-plate magnetron at four values of voltage of the other segment. Arbitrary units.

10-11. Negative-resistance Magnetron Oscillator.- Figure 10-15 shows the manner in which the current to one segment of a two-segment magnetron varies with the potential of that segment at several values of potential of the other segment in the vicinity of plate-current cutoff.[19] The current and voltage units in this figure are arbitrary. Figure 10-16 shows curves of the sum and difference of the currents to the two segments of the anode of a magnetron as a function of the difference in potential of the two segments when the potential of one segment is raised and that of the other lowered by an equal amount.[20] Since the curves of Fig. 10-15 and the difference curve of Fig. 10-16 show portions with negative slope, it follows from the discussion of Sec. 4-17 that sustained oscillations should be produced if a parallel resonant circuit is connected either in series with one anode segment and the cathode or between the two

19. McArthur and Spitzer, Proc. I.R.E., 19, 1971 Nov. (1931).
20. Kilgore, G.R., Jour. App. Physics, 8, 666 Oct. (1937).

segments. Figure 10-17 shows the circuit of the usual type of negative-resistance magnetron, in which the tank circuit is connected between the two anode segments. Because of the similarity to ordinary push-pull circuits, this arrangement is commonly called the push-pull connection.

A physical explanation for the production of negative resistance can be gained from Fig. 10-18, which shows typical electron paths, estimated from the electrostatic and magnetic fields when the magnetic field is approximately 1.5 times the cutoff value. The path in Fig. 10-18a is for equal potentials on the two segments. The electrons travel in symmetrical paths and very few reach the anode. Figures 10-18b and c show, however, that when the potential of one sector is raised and that of the other lowered, electrons travel in complicated paths and strike the sector having the lower potential. Hence if the potential of sector A exceeds that of sector B, plate current flows to sector B, and $I_A - I_B$ is negative when $E_A - E_B$ is positive.

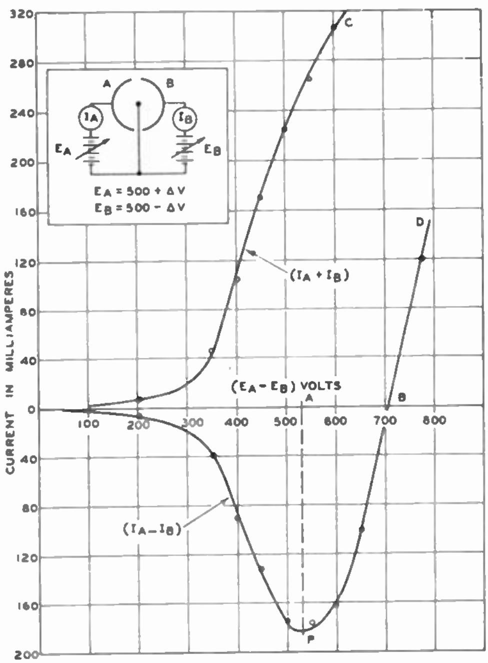


FIG. 10-16. Static characteristics of a two-segment magnetron showing negative resistance.

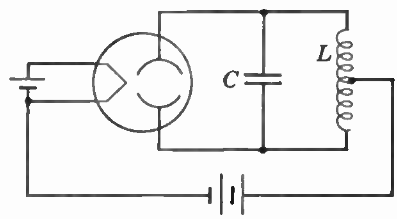


FIG. 10-17. Circuit of negative resistance magnetron oscillator.

Figure 10-18d shows an experimental verification of Fig. 10-18b obtained by photographing the ionized path of electrons in a gas-filled tube.

The efficiency and power output of the negative-resistance magnetron oscillator are limited by the same factors as negative-grid oscillators, listed in Sec. 10-1. The circuit requirements are more stringent, however, since the resonant impedance of the oscillating circuit must be higher for negative-resistance

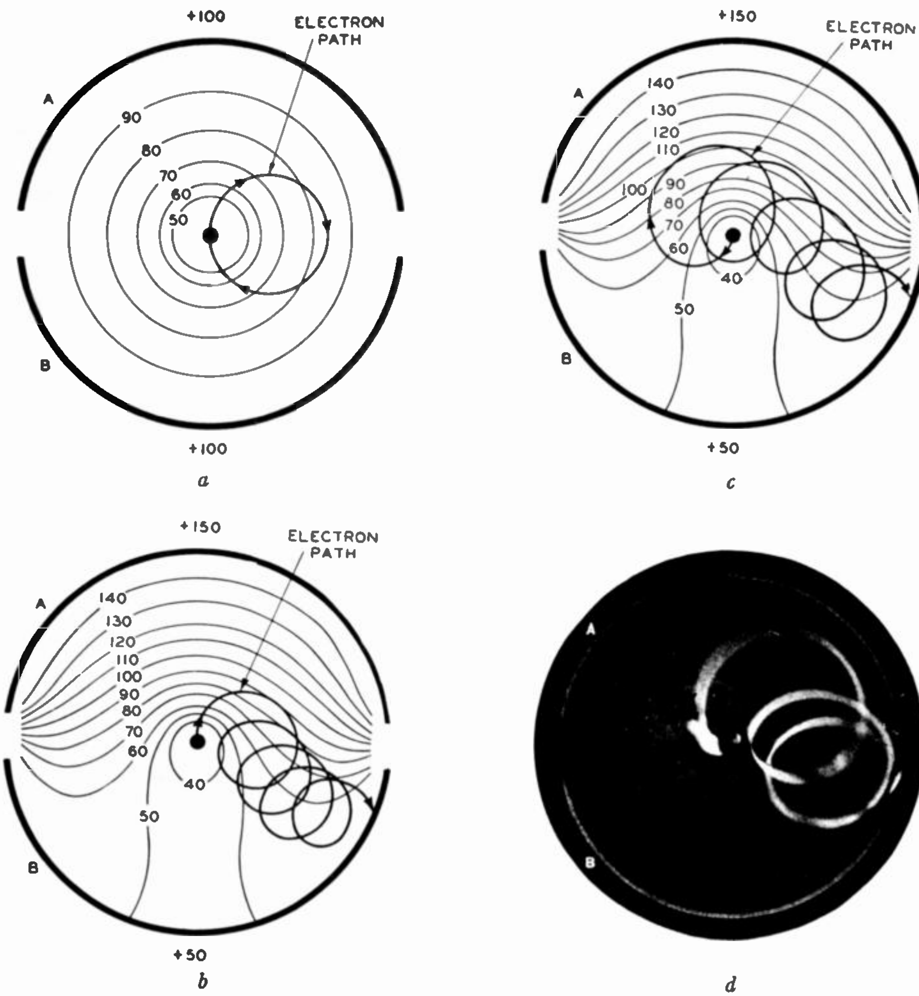


FIG. 10-18. Typical electron paths in a two-segment magnetron, showing how electrons arrive at the plate-half of lower potential. (Kilgore.)

magnetron oscillators than for negative-grid oscillators. A circuit with given losses, therefore, will not oscillate at so high a frequency in the negative-resistance magnetron oscillator, and the efficiency at a given frequency is lower than in the negative-grid oscillator. Furthermore, as indicated by eq. (10-9), in order to make the transit time small, it is necessary to use high magnetic flux density. An efficiency of 30 per cent at a wavelength of 10 cm would require a flux density of approximately 8000 gauss. The difficulty of obtaining the required high flux density is one of the limitations involved in the use of negative-resistance magnetron oscillators. Moreover, eq. (10-12) indicates that in order to operate in the vicinity of plate-current cutoff at high values of flux density, it is necessary to use high anode voltage and small anode diameter. The difficulty of obtaining adequate plate dissipation is, therefore, another serious limitation in the use of negative-resistance magnetrons at ultra-high frequencies. Plate dissipation may be increased by the use of a heavy-walled plate structure and by placing the oscillating circuit within the tube envelope. The conductors are then made of large cross section and high thermal conductivity, so that heat is dissipated by the entire circuit. Water cooling has also been used. So far, however, the use of this type of magnetron has, for the most part, been restricted to wavelengths of 50 cm and greater, although by the use of water-cooling and continuous evacuation it has been possible to obtain an output of 80 watts at 19 cm.

10-12. Transit-time Magnetron Oscillators.- The theory of operation of transit-time magnetron oscillators is in some respects similar to that of positive grid oscillators. The mechanism by which energy-absorbing electrons are removed from the interelectrode space is, however, different.

Although the magnetron with plane parallel electrodes is seldom used, the analysis of this type is simpler than that of cylindrical types, and therefore of value as an introduction. In the operation of all transit-time magnetrons the flux density is adjusted so that plate current is approximately at cutoff. Electrons then just graze the surface of the plate, and return to the cathode. The path of electrons is as shown by curve b of Fig. 10-13. Suppose that a small alternating voltage of period equal to the time taken for an electron to move to the plate and back is superimposed upon the steady plate voltage. If an electron leaves the cathode (or virtual cathode) at the instant the alternating voltage is zero and increasing in the positive direction, the electric field throughout its motion toward the plate is greater than when the alternating voltage is zero. The electron will therefore approach closer to the plate, and may strike it. If it does not strike the plate, the electron returns to the cathode. Because the difference in potential between plate and cathode throughout its return trip is less than during its outward trip, it loses less kinetic energy than it gained, and will therefore strike the cathode with appreciable velocity and will be lost to the space. The excess energy with which it strikes the cathode is derived from the source of alternating voltage. An electron that leaves the cathode a half cycle later, on the other hand, does not approach so close to the plate as when the plate voltage is constant. Since the potential difference during its return trip is greater than during its outward trip, it loses more kinetic energy in moving a given distance normal to the electrodes during the return trip, and so comes to rest before it reaches the cathode. Since it would still have had energy at this point if the anode voltage had been constant, it must have delivered energy to the source of alternating voltage. The electron can continue to oscillate back and forth with decreasing amplitude and to deliver energy to the source of alternating plate voltage until it comes to rest at a point between the anode and the cathode or leaves the interelectrode space as the result of its sideward motion. In a similar manner, all electrons that leave the cathode during the portion of the cycle when the plate voltage is rising take energy from the source of alternating voltage, but strike the plate at the end of the first outward trip or the cathode at the end of the first return trip, and so cannot make more than one round trip. All electrons that leave while the plate voltage is decreasing, however, may make many excursions, during each of which they deliver energy to the source of alternating plate voltage. Since the rate at which electrons leave the cathode on the average is constant, the source of alternating voltage gains more energy than it loses. If the alternating voltage is supplied by an oscillatory circuit, sustained oscillations are produced if the energy gained by the circuit as the result of the motion of electrons exceeds the energy lost in the circuit. An analysis similar to that made in Sec. 10-5 (Figs. 10-9 and 10-10) shows that the motion of the electrons results in the production of an effective negative resistance between the plate and the cathode.

The behavior of electrons in a cylindrical magnetron oscillator is similar to that in the parallel-plate magnetron oscillator. Without alternating plate voltage the paths are of the form of curve c of Fig. 10-14. Electrons that take energy from the source of alternating voltage strike the plate at the end of the first outward trip, or the cathode upon their return. Electrons that deliver energy to the source of alternating voltage, however, make many oscillations of decreasing amplitude, during which they precess about the filament as shown in Fig. 10-19.

In each successive cycle of oscillation of an electron that gives up energy, the distance through which it moves decreases, and the transit time goes down. The oscillation of the electron therefore advances in phase relative to the alternating plate voltage. The shift in phase eventually becomes great enough so that the electron no longer delivers energy to the

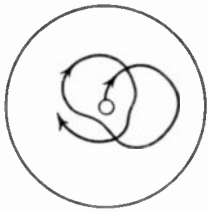


FIG. 10-19. Motion of an electron that delivers energy to a source of alternating plate voltage.

source of alternating voltage, and may start absorbing energy.[21] If the electrons remain in the interelectrode space long enough, they will absorb as much or more energy than they delivered to the source of voltage. In the parallel-plate magnetron the sideward motion of the electrons carries them out of the interelectrode space before the phase shift becomes great enough to be objectionable. In the cylindrical magnetron, however, the electrons continue to precess about the filament unless some means is provided for removing electrons before they start absorbing energy. One method of removing the electrons is by tilting the magnetic field at an angle to the filament, so that the electrons have a component of motion parallel to the filament. The optimum angle is quite critical, since it is necessary to remove the electrons before they start absorbing energy, but not before they have delivered at least a large portion of their energy to the source of alternating voltage. The angle must be changed with plate voltage. The electrons may also be removed by means of end plates, shown in Fig. 10-20, which are maintained at a potential somewhat lower than that of the anode. Since the best end-plate voltage is found to be proportional to anode voltage, the correct voltage may be readily maintained by the use of a voltage divider across the source of anode voltage. It should be noted that end plates must be used even if the field is tilted, since otherwise the electrons that leave the interelectrode space strike the tube wall and may cause sufficient heating to melt the glass. When the field is tilted, the end plates may be connected to the filament. Brillouin has shown mathematically that the cylindrical magnetron acts like a negative resistance when the frequency of the alternating plate voltage is equal to or nearly equal to the transit frequency f_t at which electrons move from the cathode to the vicinity of the

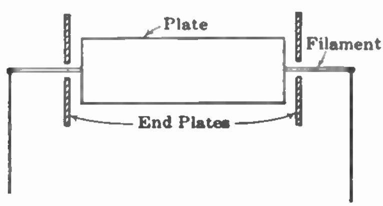


FIG. 10-20. Electrode structure of cylindrical magnetron with end plates.

21. A similar effect would result without this shift in phase, since an electron that remained in the interelectrode space would eventually come to rest and would then start oscillating in opposite phase with increasing amplitude and thus absorb energy from the source of alternating voltage.

plate and back. At cutoff, this frequency is equal to $\sqrt{2}$ times the frequency at which the electrons precess about the filament. By application of eq. (10-9) the transit frequency is

$$f_t = \frac{\sqrt{2}}{2\pi} \times 8.84 \times 10^6 \, B = 1.98 \times 10^6 \, B \quad \text{cps} \qquad (10\text{-}13)$$

Sustained oscillations may take place when the frequency of the resonant circuit has values equal or nearly equal to f_t. The corresponding wavelength is

$$\lambda_t = c/f_t = 3 \times 10^{10}/1.98 \times 10^6 \, B = 15{,}160/B \quad \text{cm} \qquad (10\text{-}14)$$

Brillouin has also shown that the tube may act like a negative resistance in other ranges of frequency above and below f_t. Sustained oscillations may take place when the frequency of oscillation of the circuit lies within these ranges if the circuit losses are not too great. The magnitudes of negative resistance in these ranges are, however, such that the criterion for sustained oscillation is not likely to be satisfied in practical circuits. The observed frequency of oscillation is therefore ordinarily that given by eq. (10-13).

10-13. Split-plate Transit-time Magnetron.- The action of the segmented-plate magnetron in producing transit-time oscillations differs somewhat from that of the cylindrical plate type, although the mechanism of energy transfer is similar. Figure 10-14d shows the motion of electrons when the plate voltages are equal and constant and the magnetic field is appreciably above the cutoff value. In addition to oscillating radially, the electrons precess about the filament with an angular velocity given by eq. (10-9). The electric field resulting from a difference in potential between the two anode segments is of the form shown in Fig. 10-21. It can be seen that this field may either increase or decrease the angular velocity of an electron, depending upon the angular position of the electron. If an alternating voltage of period equal to the time taken for an electron to move once around the filament is impressed between the anode segments, the direction of the field reverses twice in each revolution of the electron about the filament. Consequently, electrons that pass the gaps at the instants at which the field is a maximum will experience continuous angular acceleration or deceleration. Electrons that are accelerated gain energy from the source of alternating voltage.

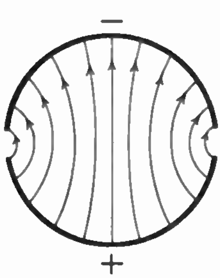


FIG. 10-21. Electric field result from potential difference between plate segments.

The increase in their angular velocity produces an increased component of force toward the filament, and so these electrons enter the filament after one excursion. Electrons that are decelerated, on the other hand, deliver energy to the source of alternating voltage. Their decrease in angular velocity reduces the average force toward the filament, and so they drift toward the plate, as shown in Fig. 10-22. Since they may make many radial oscillations before striking the plate, whereas the electrons that gain energy can make only one excursion, the source of alternating voltage gains energy.[22] Sustained oscillation may take place if the alternating voltage impressed between the plate segments is furnished by an oscillatory circuit of sufficiently low dissipation. A similar result is attained by the use of four plate segments[23] connected as in Fig. 10-23.

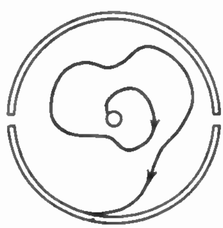


FIG. 10-22. Motion of an electron that delivers energy to a source of alternating potential between magnetron plate segments.

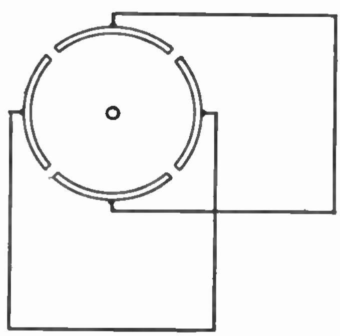


FIG. 10-23. Four-sector magnetron.

The oscillations in the two circuits are 90 degrees out of phase and the resulting voltages between the segments produce a rotating electric field. Oscillations of this type may also be produced by the use of a cylindrical anode and segmented end plates, between which the oscillatory circuits are connected.[24]

22. Experiments by Linder, Proc. I.R.E. 26, 346 March (1938), have shown that electrons average 80 to 100 excursions.
23. Yagi, H., Proc. I.R.E. 16, 715 (1928). For additional bibliography see G.C. Kilgore, Jour. App. Physics, 8, 666 Oct. (1937).
24. Okabe, K., Proc. I.R.E., 27, 24 Jan. (1939).

10-14. Practical Considerations in the Design and Use of Magnetrons.- Equation (10-13) shows that the flux density required for the operation of a transit-time magnetron increases with frequency. Since the tube must be operated in the vicinity of plate-current cutoff, it follows from eq. (10-12) that a high ratio of plate voltage to plate diameter must be used at high frequencies. The optimum plate diameter has been found to be about 1/20 of the wavelength, and the useful frequency range for a given diameter about 2 to 1. In order to minimize circuit losses and to make possible a high circuit frequency, interelectrode capacitances must be kept small by the use of short plate length. Limitation of plate area reduces plate dissipation, however, and thereby limits the power that can be generated. Great improvement in this respect can be obtained by the use of the "anode tank circuit" structure[25] illustrated in Fig. 10-24. In this type of tube the two segments of the plate are joined at one end and hence form a portion of a Lecher wire system, closed at both ends, of length $\lambda/2$. In this manner lumped capacitance is avoided. Linder has found that when the load impedance is properly chosen, interelectrode capacitance has no effect upon the frequency of oscillation of an anode-tank-circuit magnetron. He reported that a tube with an anode as large as 1.7 cm in diameter and 2.3 cm long was made to oscillate at 9 cm, and that an output of 20 watts at an efficiency of 22 per cent is obtainable at a wavelength of 8 cm.

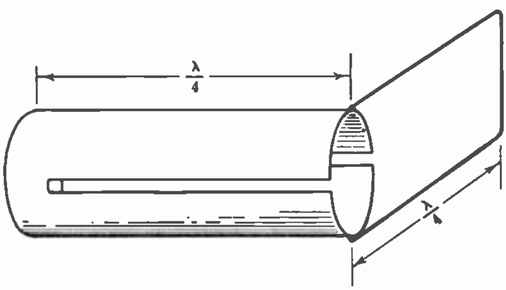


FIG. 10-24. Plate structure of anode-tank-circuit magnetron.

Figure 10-25 show the design of a typical transit-time magnetron with positive end plates. It should be noted that the oscillating circuit consists of a Lecher wire system with a short-circuiting bridge at the first voltage node. In order to make the frequency high, this bridge is placed very close to the plates. The Lecher wire system as a whole must, of course, be tuned, and the output may be taken from the circuit by coupling to the Lecher wire. Transit-time magnetron oscillators will oscillate with circuits having considerably lower resonant impedance than negative-resistance magnetron oscillators, and require much lower magnetic field strength. Outputs of the order of 2.5 watts at an efficiency of 12 per cent at a wavelength of 9 cm were reported by Linder in 1936. Tubes of more recent design are capable of delivering considerably higher outputs.

25. Linder, E. G. Proc. I.R.E., 27, 732 Nov. (1939).

Transit-time magnetrons have been made to oscillate at wavelengths as short as 6 millimeters.[26]

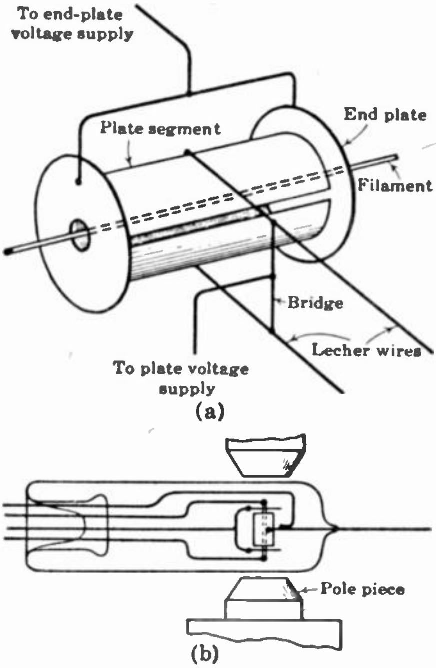


FIG. 10-25. (a) Structure of end plate magnetron. (b) Method of applying magnetic field.

In order to obtain stable oscillations of constant frequency with transit-time magnetrons it is necessary to use well-regulated voltage supplies and very constant magnetic field. Because small changes in filament current can start and stop oscillation, probably because of changes in space charge, the filament current must also be maintained constant. Unfortunately the literature appears to abound with an absence of design information for magnets suitable for use with magnetrons. One difficulty likely to be encountered is the change of resistance of the field coils as they warm up, with resulting change in current. This difficulty can be avoided by the use of electromagnets designed to operate at high voltage and low current. By means of a current regulator of the type described in Sec. 2-22, it is then possible to maintain constant flux density. Permanent magnets are now available that are capable of providing the required flux, which may be adjusted by means of a magnetic shunt. It has been found possible to obtain sufficient frequency stability so that an audible beat note is obtained between two oscillators operating above 1000 megacycles.[27]

Another problem encountered in the use of transit-time magnetrons results from a cumulative increase of plate current and rise in filament and plate temperature. This phenomenon undoubtedly results from the bombardment of the cathode by electrons that return to the cathode with excess energy. There

26. Cleeton, C.E., and N.H. Williams, Phys. Rev., 40, 234 (1934); Cleeton, C.E., Physics, 6, 207 (1935); Cleeton, C.E. and Williams, N.H., Phys. Rev. 50, 1091 (1936).

27. Megaw, E.C.S., G.E.C. (London) vol. 7, May (1936).

appear to be two probable ways in which electrons can acquire excess energy. The first discussed in Sec. 10-10, is the conversion of some of the energy of orbital motion into energy of random motion as the result of collisions between electrons. This effect is present even when the tube is not oscillating. The second is the transfer of energy from the source of alternating plate voltage to those electrons which leave the cathode (or virtual cathode) at the wrong time in the cycle of alternating voltage. This effect is present in all transit-time magnetron oscillators. The rise in filament temperature may increase the number of electrons moving in the interelectrode space and thus increase the plate current and also result in further bombardment of the cathode and increase of filament temperature. Unfortunately, it is found that this action is most likely to occur when the circuit is adjusted so as to give highest output. Once started, the cumulative action takes place so rapidly that, unless adequate protective measures are taken, the filament burns out before any circuit adjustments can be made. For this reason it is necessary to make use of a control circuit that lowers either the filament current or the plate voltage when the plate current rises. A control circuit that lowers the anode voltage was discussed in Sec. 2-22. Figure 10-26 shows a thyratron phase-control circuit that lowers the filament current.

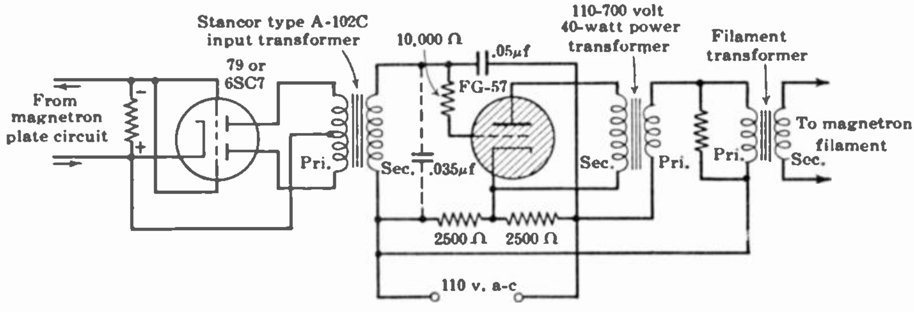


FIG. 10-26. Filament-current stabilizer for magnetron.

10.15. Cavity Resonators.- In Sec. 10-2 it was pointed out that the efficiency of an ultra-high-frequency oscillator using a Lecher wire system can be raised and the upper frequency limit increased by placing the tube elements at the center of a half-wave Lecher wire resonator instead of at the end of a quarter-wave resonator. Because the two halves of the half-wave system are in effect in parallel, the RI^2 losses due to charging currents are reduced. This suggests that considerable reduction in dissipation should be effected by using a large number of quarter-wave Lecher wires in parallel. The various loops may be arranged over the surface of revolution generated by rotation of a quarter-wave loop about an axis in the plane of the loop at the open end, as shown in Fig. 10-27. If the number of loops is increased sufficiently, the circuit eventually becomes a closed chamber the boundaries of which conform to the surface of revolution generated by rotation of one loop.[28] The electromagnetic fields are then confined to the inside of the chamber. This is a distinct advantage, since it prevents energy loss caused by radiation, and changes of resonance frequency caused by body capacitance. Because it is usually not convenient to place the elements of an electron tube within such a cavity resonator, confinement of the electromagnetic fields to the inside of the cavity makes it necessary to employ special methods of delivering and removing energy from the resonator. This may be accomplished by shooting electrons through the resonator at a point where the electric field strength is high, by inserting a probe (small antenna) into the chamber at a point where the electric field strength is high, or by introducing a small conducting loop into the chamber at a point where the magnetic field strength is high. The probe or loop is connected to a concentric cable with the aid of a suitable impedance-matching device.

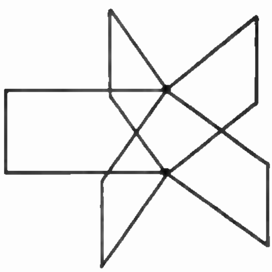


Fig. 10-27. Quarter-wave resonators connected in parallel.

The subject of cavity resonators may also be approached from the point-of-view of wave guides, to be discussed in Ch. 14. In that chapter it will be shown that electromagnetic radiation may be propagated through a conducting pipe. If one end of such a wave guide is closed by a wall of high conductivity, the electromagnetic waves are reflected by this wall. The phase of the electric field is reversed on reflection, whereas the phase of the magnetic field is not changed. If the conductivity of the wall is high, the amplitudes of the incident and reflected waves are nearly equal. Since the incident and reflected waves travel with the same velocity, the electric fields of the direct and reflected waves reinforce each other at distances from the reflector equal to $(2n - 1)\lambda_g/4$, where n is an integer and λ_g is the wavelength in the guide, and cancel at distances equal to $n\lambda_g/2$. There are, therefore, loops and nodes of electric field spaced a quarter of a wavelength apart. Inasmuch as there is negligible

28. This introductory picture of transition from wire circuits to resonators is meant for illustrative purposes only and is not intended to be rigorous. The reader will recognize obvious deficiencies, such as ignoring mutual impedance between the wire circuits, etc.

potential difference in the reflector if its conductivity is high, the electric field is also negligible at the reflecting surface. The magnetic field, on the other hand, has maximum amplitude at the reflecting surface and at distances equal to $n\lambda_g/2$ from the reflector, and minimum amplitude at distances $(2n - 1)\lambda_g/4$. A second reflecting surface may be added at any nodal point of the electric field without changing the general form of the standing waves. Then, if the point at which the energy is introduced into the guide lies between the two reflecting surfaces, standing waves of constant amplitude will continue to exist. Apparently, the shortest resonator that can be made in this manner will have a length equal to one-half the wavelength within the guide. Since the losses increase with guide length, this half-wave resonator will have lower loss than any whose length is equal to a larger number of half wavelengths. Because the velocity of propagation of radiation within the guide is higher than that in free space, the wavelength within the guide is greater than that in free space. A resonator whose length may be arbitrarily small compared to the free space wavelength may be obtained by operating a closed guide at one of its critical frequencies.

In general, any closed chamber whose walls have high conductivity may serve as a cavity resonator. Because the wave energy may be propagated in more than one direction and because, as already pointed out, the standing waves may have one or more loops, such resonators may in general oscillate in a large number of "modes." Cavity resonators suitable for use in the generation of oscillations are of relatively simple form, and are usually used in the lowest-frequency, or fundamental, mode of oscillation. Although they have been used in other types of oscillators, cavity resonators have found their principle application in "velocity-modulation" tubes such as the klystron, to be described in a succeeding section.

Figure 10-28 shows the forms and distribution of the electric and magnetic fields within a cylindrical resonator oscillating in its fundamental mode with the electric field E parallel to the axis. Electric currents flow in the wall of the cylinder as shown, and charges are stored in the top and bottom faces. If the wall were a perfect conductor, the currents would flow entirely on the inner surface of the wall, and there would be no penetration into the wall. With finite conductivity, however, the energy and currents may be shown to penetrate into the wall.

In computations of energy loss a useful factor is the so-called "skin depth," δ, which is equal to $\sqrt{\rho/f}\,/2\pi$, where ρ is the resistivity and f is the frequency. It should be noted that the distance to which currents penetrate into the conductor is not sharply defined and though it is often measured by δ,

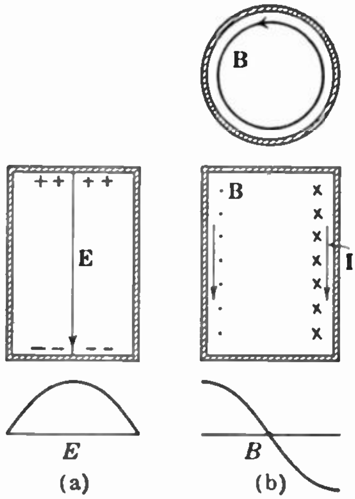


FIG. 10-28. Charge, currents, and field distribution in cylindrical cavity resonator (a) at instant of maximum electric field, and (b) one-quarter cycle later.

serious errors may sometimes result from assuming that the skin depth is the depth of current penetration.

10-16. Characteristics of Cavity Resonators.[29] - The characteristics of a parallel resonant circuit with lumped constants are fully determined by the three lumped parameters, the inductance L, the capacitance C, and the resistance R in series with the inductance. Since cavity resonators do not have lumped parameters, it is necessary to set up somewhat arbitrary definitions of equivalent parameters in order to determine the relative merits of cavity resonators of various shapes and to compare them with conventional resonant circuits. Although it is possible to define equivalent values of inductance, capacitance, and resistance, it is more useful to define, instead, three quantities that are important in specifying the performance of conventional resonant circuits. These three quantities are the resonance frequency, $1/2\pi\sqrt{LC}$; the circuit Q, $\omega L/R$, which is indicative of the selectivity of the circuit; and the anti-resonance, or equivalent shunt impedance, L/RC, which enters into the criterion for oscillation (see, for instance, Sec. 4-19). Since these three quantities all involve the three basic circuit parameters, the circuit is specified as completely by these quantities as by the basic circuit parameters. It is desirable to specify the characteristics of a cavity resonator by means of analagous quantities.

It is shown in Ch. 1 that the Q of a conventional parallel resonant circuit, $\omega L/R$, is equal to 2π times the ratio of the peak value of the energy stored in the electric or magnetic field to the energy dissipated per half cycle at resonance. The ratio of the energy stored in a cavity resonator to the energy lost per half cycle is a suitable indication of the equivalent Q. Since the electric field is zero at the instant when the magnetic field is a maximum, the stored energy may be calculated as the energy of the magnetic field alone within the cavity at the instant this field is a maximum.

29. Hansen, W.W., Jour. App. Physics, 9, 654 Oct. (1938);
Hansen, W.W., and Richtmyer, R.D., Jour. App. Physics, 10, 189 March (1939).

The energy lost may be assumed to be equal to the energy that penetrates[30] into the "skin." The energy in emu stored in a magnetic field is equal to $\mu/8\pi$ times the integral of B^2 throughout the space in which the field exists. On the basis of these facts and the further fact that $\delta \ll \lambda$, Hansen has defined Q for a cavity resonator by the following equation:

$$Q = \frac{\int_{vol.} B^2 \, d\tau}{\delta \int_{surf.} B^2 dS} = \frac{\int_{vol.} B^2 \, d\tau}{\lambda \int_{surf.} B^2 dS} \cdot \frac{\lambda}{\delta} \qquad (10\text{-}15)$$

in which $d\tau$ is an element of volume, dS an element of the surface of the cavity, λ the wavelength in free space, δ the skin depth, B the magnetic flux density at the instant at which the electric field intensity E is zero. The product δdS may be considered an element of volume of the conductor at its surface. The numerator is an integral throughout the volume of the cavity and the denominator integral is an integral over the surface of the conductor bounding the cavity.

The quantity

$$\frac{\int_{vol.} B^2 \, d\tau}{\int_{surf.} B^2 dS} \qquad (10\text{-}16)$$

is substantially the ratio of a volume to the surface enclosing it and is therefore proportional to a linear dimension of the cavity. For example, if the radius a of the circular cylinder shown in Table 10-I were doubled, the volume of the cavity would be quadrupled, the surface of the enclosing area doubled (except for the surfaces at the ends of the cylinder) and, speaking roughly, (10-16) would be doubled provided B changed in such a way that its effect on the two integrals was practically the same (which is an approximation to fact). The ratio (10-16) is therefore proportional to a linear dimension of the cavity. Since the wavelength to which the cavity resonates is also proportional to a linear dimension of the cavity (see Table 10-I) the first factor in the second form of eq. (10-15) is to a first approximation independent of wavelength, but characteristic of the shape of the cavity. It is evident from the form of this ratio that in order to have a high value of Q a cavity should have a high ratio of volume to surface. Since the skin depth δ is proportional to the square root of the wavelength, the second factor of eq. (10-15) shows that, for a cavity of given shape, Q varies as the square root of the wavelength. Since δ varies as the square root of the resistivity, Q also decreases with increase of resistivity.

The equivalent shunt or input impedance L/RC of a conventional parallel resonant circuit is a resistance which, when

30. This assumption is based on the thesis that most of the energy entering the conductor is dissipated, and does not return to the cavity.

divided into the square of the effective voltage across the circuit, gives the power loss of the circuit. In a cavity resonator a quantity analagous to the voltage across a parallel circuit is the line integral of the electric field along a path parallel to the electric field in the plane or line in which the field is greatest, which is equal to the time rate of change of the total magnetic flux through the loop formed by the path.[31] The equivalent shunt impedance of a cavity resonator may, therefore, be defined as the resistance which when divided into the square of this line integral gives the power dissipated in the cavity. The following expression for R_e is based upon this definition:[32]

$$\frac{\omega^2[\int_{surf.} \underline{B}\cdot d\underline{S}]^2}{R_e} = \frac{\delta}{2T} \int_{surf.} B^2 dS$$

or

$$R_e = 16\pi^2 \frac{[\int_{surf.} \underline{B}.d\underline{S}]^2}{\lambda^2 \int_{surf.} B^2 dS} \cdot \frac{\lambda c}{\delta} \qquad (10\text{-}17)$$

in which c is the velocity of light. (The dot in the integrand in the numerator of eq. (10-17) indicates that the element of surface is multiplied by the component of flux density $\underline{B}$ normal to the surface.) Table 10-I lists formulas for λ, Q, and R_e for five types of cavities, and numerical values of Q and R_e for copper resonators resonant at a fundamental wavelength of 10 cm.

It should be noted that the harmonic frequencies at which cavities resonate do not in general have integral ratios to the fundamental frequencies. It is probably also well to mention that formulas for λ, Q, and R_e can be derived only for cavities of relatively simple shape. These quantities can, however, be determined experimentally for more complicated shapes.

10-17. Velocity-modulation Tubes.- The operation of all transit-time tubes is based upon the transfer of energy from moving electrons to a source of alternating voltage by the action of an electric field produced by the alternating voltage. In order for such a tube to be capable of acting as an amplifier or oscillator, it is necessary that the energy delivered by the electrons shall on the average exceed the energy absorbed by electrons from the alternating voltage source. Since electrons depart from an electron emitter at random intervals, the electrons in a normal space current or beam of cathode rays are just as likely to pass through the field at such a time that they gain energy as at such a time that they lose energy. Hence some special means must be provided in order to make the energy

31. See Ch. 12, where this and many other relations used in this section are developed.

32. Hansen and Richtmyer, loc. cit. The integral in [·] in (10-17) is taken over only the surface on one side of the loop.

delivered to the field exceed the energy taken from the field. In Sections 10-5 and 10-12 it was shown that this is accomplished in the positive-grid oscillator and the magnetron by removing from the space all electrons that gain energy. The mechanism is such, however, that the electrons must gain energy before they can be removed from the space. The increase of energy of electrons before removal from the space is undesirable for a number of reasons, the most obvious of which is the reduction of efficiency of the device. The energy gained by the electrons before their removal appears at one or more electrodes and causes a rise in temperature which, in magnetrons, may become cumulative and which limits the power that can be delivered by the tube. Since such a device can cause sustained oscillation only when the power delivered to the oscillating circuit exceeds the power dissipated by the circuit, the fact that some electrons take energy from the oscillating circuit increases the required Q of the oscillatory circuit and reduces the frequency limit of oscillation attainable with a given circuit. Furthermore, the fact that no more than 50 per cent of the electrons emitted from the cathode can deliver energy to the oscillatory circuit or other source of alternating voltage means that the emission current is used inefficiently. These objectionable effects are minimized in velocity-modulation tubes by making most of the electrons pass through the electric field at such times that they deliver energy to the source of the alternating electric field. This is accomplished by speeding up the electrons that would normally pass through the field too late and slowing down the electrons that would normally pass through the field too soon.

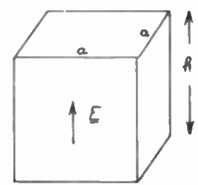
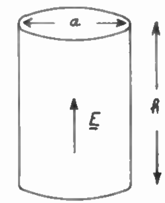
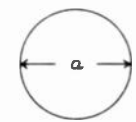
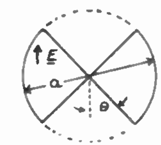
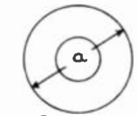
10-18. The Klystron.- One type of velocity-modulation tube is the klystron, developed by the Varian brothers.[33] In this tube the velocity of the electrons is changed in the desired manner by sending them through a cavity resonator in a direction parallel to the electric field. If the time taken for an electron to pass through the resonator is too great, the field may reverse while the electron passes through and thus result in less change of velocity than if the transit time were shorter. For this reason the time taken for an electron to pass through the resonator is made appreciably less than half the period of oscillation of the resonator. Although it is possible to make h (see Table 10-I) small in a cavity resonator in the shape of a square prism or a circular cylinder, and thus to make the path along the electric field short, the resulting low ratio of volume to area makes Q too low. Higher values of Q can be obtained by the use of resonators of the shapes shown in the fifth column of Table 10-I or of the shape shown in Fig. 10-29, which is a

33. Varian, R.H., and Varian, S.F., Jour. App. Physics, 10, 321 May (1939).

Table 10-I[1]

Characteristics of Various Types of Resonators

(δ is skin depth. For Copper and a wavelength of 10 cm., $\lambda/\delta = 8.32 \times 10^4$.)

	Square Prism	Circular Cylinder	Sphere	Sphere and Cones (Hour glass)	
				 Cross-section view	 Top view  Cross-section view
Wavelength λ	1.41a	1.31a	1.14a	2a	approximately 1.2a
Q	$\frac{0.353\lambda}{\delta} \frac{1}{1+\frac{a}{2h}}$	$\frac{0.383\lambda}{\delta} \frac{1}{1+\frac{a}{2h}}$	$0.318\frac{\lambda}{\delta}$	$0.1095\frac{\lambda}{\delta}$ (θ = optimum value = 34°)	$0.04\frac{\lambda}{\delta}$
Equivalent shunt resistance R_e	$170\frac{h}{\delta} \frac{1}{1+\frac{a}{2h}}$	$185\frac{h}{\delta} \frac{1}{1+\frac{a}{2h}}$	$104\frac{\lambda}{\delta}$	$32\frac{\lambda}{\delta}$ (θ = optimum value = 9°)	$3\frac{\lambda}{\delta}$
Q at 10 cm. (Copper walls; a = h)	24,100	26,200	26,500	9,120	3,000
R_e in ohms at 10 cm. (Copper walls; a = h)	6.67×10^6	7.88×10^6	8.63×10^6	2.66×10^6	2.5×10^5

1. Data listed in columns 1 through 5 of this table are taken from Fansen and Hansen and Richtmyer, loc. cit. Data listed in column 6 was presented by Hansen in a lecture.

modification of the shape shown in the fourth column of Table 10-I. The constricted portion of the resonator is made in the form of grids in order that electrons may enter and leave the resonator. To minimize losses, the conductivity of those grids must be high. The change in velocity of the electrons in passing through the resonator causes them to arrive at certain points beyond in groups or "bunches." For this reason the resonator is called a "buncher." If the electrons are allowed to pass through a second cavity resonator so placed that the electrons pass through in groups, and if the resonator oscillates in such phase that the field retards the electrons, energy will be delivered to this resonator. This energy will be greatest if the electric field is a maximum when the center of the group of electrons passes through the center of the resonator and if the field is just strong enough to bring the slowest electrons to rest at the farther wall of the resonator. If the field is stronger than this, some electrons will acquire a velocity in the reverse direction within the resonator and will thus again take energy from the resonator. The second resonator is called the "catcher." Electrons that pass through the catcher may strike a collector from which they are returned to the cathode through an external circuit. In some types of klystrons, electrons that emerge from the catcher pass through a grid into a field-free chamber formed by an "end cap" attached to the catcher. The essential structure of collector and end-cap type klystrons is shown in Figs. 10-30a and 10-30b. The first grid, which is operated at a relatively low positive voltage, serves both as a control electrode to vary the beam current, and as a focusing electrode.

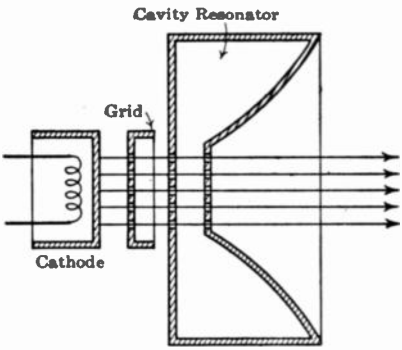


Fig. 10-29. Type of cavity resonator (rhumbatron) used in klystron.

10-19. Theory of Electron Bunching.- The manner in which the buncher produces groups of electrons can be understood by considering the motion of individual electrons. An electron that passes the center of the buncher, just as the electric field is changing from retarding to accelerating, leaves the buncher with the same velocity with which it entered. An electron that passed the center a few electrical degrees earlier leaves with reduced speed, whereas one that passes a few electrical degrees later leaves with increased speed. If the space beyond the buncher is field-free, the differences of speed cause the electrons ahead of and behind the one with unchanged speed to draw nearer to it.

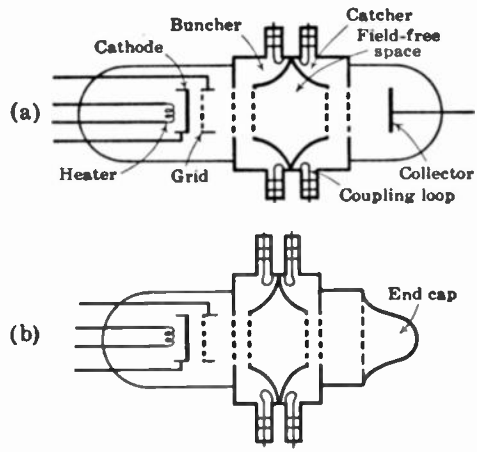


FIG. 10-30. Essential structure of klystrons

Another electron of unchanged speed that leaves the buncher a half cycle later has its neighbors draw away from it. Consequently, at a suitable distance from the third grid the stream contains groups of electrons denser than the stream at the entrance to the buncher, separated by regions of lower density. The action of the buncher can also be shown graphically by means of the "Applegate diagram" of Fig. 10-31. In this diagram the lines indicate the positions of individual electrons as a function of time, the slopes being proportional to the electron velocities. The buncher is assumed to be of

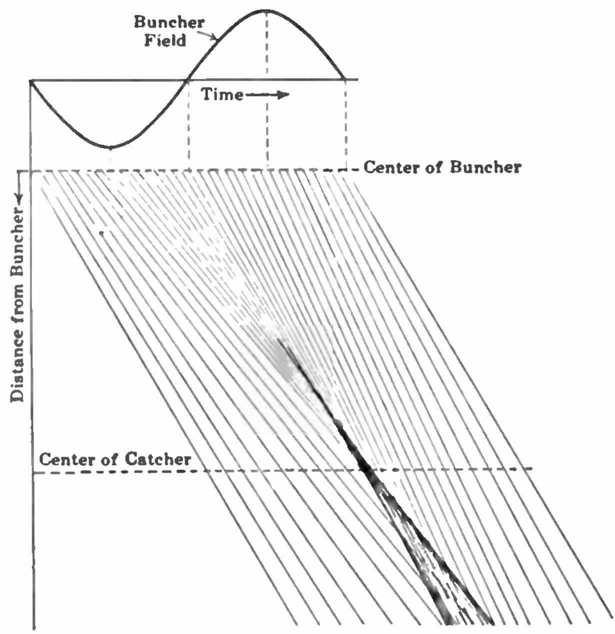


FIG. 10-31. Applegate Diagram.

negligible length. The lines are drawn under the assumption that the electron velocities are all equal before the electrons pass through the buncher, that the electrons pass the buncher at equal intervals, and that the change in velocity of an electron in the buncher is $v_1 \sin 2\pi f t_1$ where v_1 is the greatest change in velocity experienced by any electron, f is the frequency of the alternating field in the buncher and t_1 is the time at which the electron passes the center of the buncher.[34] It can be seen that along the dotted line indicating the position of the catcher the lines are close together once in each cycle and widely separated at intermediate times. This indicates that most of the electrons arrive at the catcher in groups. Extension of the lines of Fig. 10-31 shows that there are other positions, farther from the buncher, at which electrons arrive in groups. It should be pointed out that Fig. 10-31 neglects the mutual repulsion of electrons, which tends to produce "debunching" or spreading apart of the electrons in the bunches.

In practice, the change in velocity of electrons in the buncher is small in comparison with the initial velocity v_0 given the electrons by the direct voltage impressed between the cathode and the buncher. Hence, practically equal numbers of electrons approach and leave the buncher at all times, and currents induced in the buncher by approaching electrons are offset by currents of opposite sign induced by departing electrons. For this reason the external circuit current to the buncher and the power loss in the buncher are small. The small power loss may also be explained by the fact that energy given to electrons in one-half of the cycle is offset by an equal amount of energy taken from electrons during the other half cycle.

A mathematical analysis of bunching has been made by Webster.[35] In the course of this analysis the assumption is made that all electrons leaving the buncher in a certain time interval will arrive at the catcher during another time interval and that no other electrons will arrive at the catcher during the latter interval. Because of the fact that the more rapidly moving electrons may overtake and pass the more slowly moving electrons, this assumption does not appear to be justifiable, particularly since it leads to an expression for electron current at the catcher which may be negative during portions of the cycle when the distance s to the catcher is large. This analysis, a graphical analysis based on Fig. 10-31, and experiments on klystrons indicate, however, that at a certain distance from the

34. Actually the kinetic energy, rather than the velocity, varies sinusoidally. If the change in velocity is small in comparison with the initial velocity, however, the variation of velocity is very nearly sinusoidal.

35. Webster, D.L., Jour. App. Physics, 10, 501, 864 July and Dec. (1939).

buncher the wave of current is sharply peaked, and that beyond this point the current wave is doubly peaked, and broader. Since the current wave is periodic, it may be expanded in a Fourier series. The current wave therefore contains harmonics. This fact indicates that the catcher should take energy from the electron stream if it oscillates at a harmonic of the frequency of the buncher, provided that the relative phase of the two oscillations is correct. Ordinarily the buncher and the catcher are similar and are tuned to resonate at the same frequency; but in a frequency-multiplier type of klystron the catcher is tuned to some harmonic of the buncher frequency.

In order that an electron shall give up the greatest amount of energy to the catcher, it is necessary that the electron shall arrive at the center of the catcher at the instant when the electric field at the center of the catcher has its maximum negative (retarding) value. Similarly, in order for a symmetrical group of electrons to deliver as much energy as possible, the center of the group must arrive at the center of the catcher when the catcher field has its maximum negative value. Since the electron at the center of the group is the one that passed the center of the buncher at the instant when the buncher field was changing from negative to positive, the time s/v_0 required by this electron to move through the distance s from the center of the buncher to the center of the catcher must be equal to the time interval between the zero value of the buncher field and a negative maximum of the catcher field, as shown in Fig. 10-32. From this figure it can be seen that for maximum energy transfer the catcher field must lag the buncher field by the

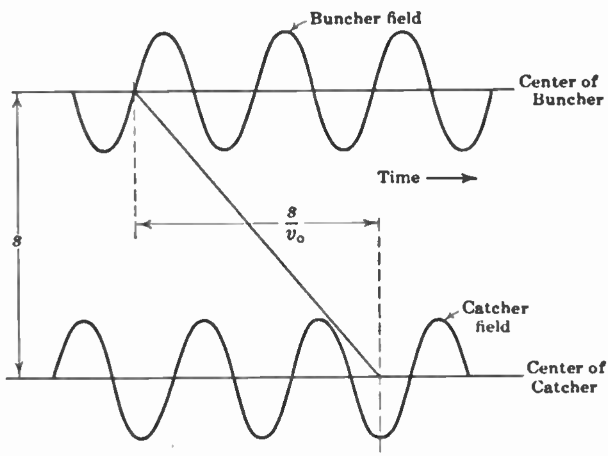


FIG. 10-32. Phase relation between buncher and catcher fields.

angle

$$\frac{s}{v_0} \times \frac{2\pi}{T} + 2\pi(\frac{1}{4} - n) \qquad (10\text{-}18a)$$

where T is the period of the field and n is any integer. The phase angle of the catcher field, relative to the buncher field is, therefore,

$$\varphi = -(\omega s/v_0 + \pi/2) + 2n\pi$$
$$= 2n\pi - (2\pi sc/\lambda v_0 + \pi/2) \qquad (10\text{-}18b)$$

in which c is the velocity of light, λ is the wavelength corresponding to the resonator frequency, and n is an integer. ($\omega = 2\pi c/\lambda$)

Up to this point nothing has been said regarding the means by which the catcher is kept in oscillation. The motion of an electron through a cavity resonator sets up an electromagnetic field within the resonator. In a klystron the electrons pass through the catcher in bunches, and so the fields resulting from individual electrons will, for the most part, reinforce each other and thus produce a resultant field of appreciable magnitude. Because the bunches pass through the catcher at time intervals equal to the natural period of oscillation of the catcher, the catcher is set into oscillation. Since the amplitude of oscillation can build up only if the catcher gains energy, the oscillation automatically tends to assume the proper phase relative to the cycle of arrival of electron bunches to result in maximum transfer of energy from the electrons to the catcher.

A series of diagrams similar to that of Fig. 10-31, constructed for various values of initial velocity v_0, frequency f, and maximum change in velocity v_1, shows that the distance from the center of the buncher at which best bunching is obtained increases with increase of v_0, with decrease of v_1, and with decrease of f. For this reason, in a klystron operated at a given frequency, an increase of accelerating voltage must be accompanied by an increase in buncher field strength for best bunching in the plane of the catcher.

Webster[36] and Condon[37] have shown that the maximum efficiency theoretically obtainable in a klystron is 58 per cent.

10-20. Debunching.- Forces of repulsion between electrons, neglected in the foregoing discussion, cause the groups of electrons to spread out, and thus reduce the efficiency of energy transfer. This effect is greatest in tubes with dense

36. Webster, D.L., Jour. App. Physics, 10, 501 July (1939).
37. Condon, E.U., Jour. App. Physics, 11, 502 July (1940).

electron beams operating at relatively low frequencies and necessitates the use of somewhat higher buncher field than would be required for best operation without debunching.[38]

10-21. The Klystron Oscillator.- In using the klystron as an oscillator, the buncher is kept in oscillation by feeding back to the buncher some of the energy delivered to the catcher by electrons. Equation (10-18) shows that the energy delivered to the catcher is greatest when the phase of oscillation of the catcher relative to the buncher has certain values. The amplitude of oscillation is therefore greatest when the angular delay of the feedback circuit has the values given by eq. (10-18a). If this angular delay time is represented by the symbol φ_f, eq. (10-18a) shows that maximum amplitude of oscillation obtains when the accelerating voltage is adjusted so that

$$\frac{sc}{v_o\lambda} = n - \frac{\varphi_f}{2\pi} - \frac{1}{4} \qquad (10\text{-}19)$$

Since $eV = 1/2\, mv_o{}^2$, in which V is the potential difference between the cathode and the buncher, $c/v_o = 500/\sqrt{V}$. Therefore

$$\frac{500s}{\lambda\sqrt{V}} \cong n - \frac{\varphi_f}{2\pi} - \frac{1}{4} \qquad (10\text{-}20)$$

Equation (10-20) indicates that there is a series of voltages at which the amplitude of oscillation is a maximum. There is, of course, a range of voltage about each of these values throughout which sufficient energy is delivered to the catcher to compensate for energy dissipated in the catcher and buncher or delivered to an external load. If the values of $500s/\lambda\sqrt{V}$ at which the amplitude is a maximum are plotted against n, n being zero for the highest voltage, 1 for the next voltage, etc., a series of points is obtained that should lie on a straight line having unit slope and an intercept on the n-axis equal to $(\varphi_f/2\pi + 1/4)$. Actually, it is usually found that the points lie on two parallel lines of unit slope. This is explained by the fact that the coupling between the two resonators exceeds the critical value. The coupled resonators can then oscillate at either of two frequencies, one of which lies above, and the other below, their natural frequency.

According to eq. (10-20), if the angular delay φ_f of the coupling network between the two resonators remains constant, the condition necessary for most efficient absorption of energy by the catcher is obtained at definite values of voltage. This would indicate that a small change of accelerating voltage would produce a relatively large change in amplitude of oscillation and might cause oscillation to cease. It may be shown, however, that a change in the frequency of oscillation results in a

38. Webster, D.L., Jour. App. Physics, 10, 864 Dec. (1939).

change in the delay angle of the coupling circuit. If the voltage is changed by a small amount, therefore, the frequency of oscillation changes sufficiently so that equation (10-20) still holds. For this reason the circuit may oscillate over a relatively wide range of voltage, but the frequency of oscillation varies with accelerating voltage. Differentiation of eq. (10-20) gives the equation

$$\frac{\partial \varphi_f}{\partial V} = \frac{500\pi s}{\lambda V^{3/2}} \tag{10-21}$$

For close coupling between the resonators, the change in φ_f with frequency at one peak of the doubly-peaked resonance curve is given by the relation

$$\frac{d\varphi_f}{df} = \frac{Q}{f} \tag{10-22}$$

Hence

$$\frac{\partial f}{\partial V} = \frac{\partial \varphi_f}{\partial V}\frac{df}{d\varphi_f} = \frac{500\pi s f}{\lambda Q V^{3/2}} \tag{10-23}$$

and

$$\frac{\Delta f}{f} = \frac{500\pi s \Delta V}{\lambda Q V^{3/2}} \tag{10-24}$$

In practical tubes the frequency variation turns out to be about 10,000 cycles per volt at 10 cm for tight coupling, and about half as much when the coupling is reduced below the critical value.

A typical diagram of connections for a klystron oscillator is shown in Fig. 10-33. Because the positive side of the

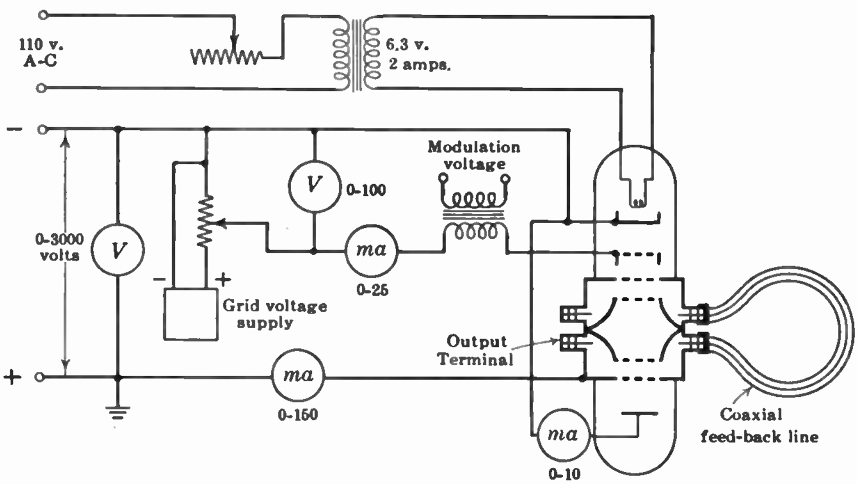


FIG. 10-33. Circuit diagram of klystron oscillator.

accelerating voltage is connected to the exposed metal portions of the tube, the positive terminal of the voltage supply is grounded, and the cathode is at high negative potential. All transformers and power supplies must be insulated for 3000 volts direct voltage. For most applications requiring an accelerating voltage of 1000 volts or less, the grid voltage may be derived from the same supply as the accelerating voltage by the use of a voltage divider. For the sake of frequency stability, the voltage supply should be well regulated. In order to prevent the flow of excessive grid current, the grid voltage should be applied after the accelerating voltage. The buncher and catcher are tuned mechanically to the desired frequency by changing the spacing between the grids (the constricted portion of the resonators through which the electrons pass). This can be accomplished by a single tuner, but before the tube will oscillate, the buncher and catcher must be tuned to the same frequency. Since oscillation is obtained over only small ranges of accelerating voltage, the values of which cannot be determined until the circuit oscillates, the initial tuning adjustment may prove to be laborious. The procedure can be greatly simplified by connecting in series with the direct accelerating voltage an alternating voltage of sufficient amplitude to insure that the resulting pulsating voltage will pass through one range in which oscillations will occur. The buncher is usually coupled to the catcher through a short length of concentric cable, as shown in Fig. 10-33.

Figure 10-34 shows the essential form and circuit of a reflex klystron oscillator, in which a single resonator acts both as the buncher and the catcher. After the electrons pass through the resonator, their direction of motion is reversed by means of a negative electrode. By proper adjustment of voltages they can be made to pass through the resonator again at the proper time so that energy is delivered to the resonator. A certain resemblance to a positive grid triode oscillator may be recognized.

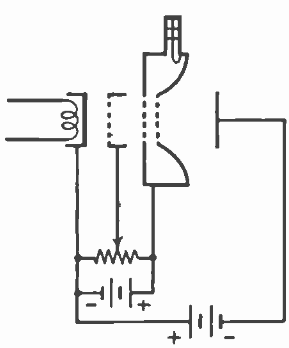


FIG. 10-34. Circuit of reflex klystron oscillator.

A typical klystron, operating at an accelerating voltage of 3000 volts is capable of delivering an output of 10 watts at wavelengths ranging from 9.5 to 11.5 cm.

10-22. Other Uses of Klystrons.- If the buncher of a klystron is driven by an external source of power of the proper frequency, and the electron beam current is high enough to deliver more energy to the catcher than is required to drive the buncher, the klystron functions as an amplifier. About 15 per cent efficiency may be obtained in klystron power amplifiers, and a voltage amplification of 20 per stage at 10 cm is possible in voltage amplifiers. If the buncher is driven both by an external source and by additional power fed back from the catcher, the klystron acts as a regenerative amplifier.

When the klystron is provided with a grid between the catcher and the collector, the tube may be used as a detector. The grid is maintained slightly positive so that when no excitation is applied to the buncher most of the electrons emerging from the catcher strike the collector. Excitation of the buncher results in the building up of a field in the catcher, which slows down some of the electrons and thus reduces the collector current. The bias of the auxiliary grid can also be adjusted so that the collector current is increased when the buncher is excited. Any other mechanism for sorting electrons according to velocity, such as a magnetic field at right angles to the electron beam, may be used in place of the auxiliary grid.

As indicated in Fig. 10-33, the grid between cathode and buncher may serve as a modulator. Because of the variation of frequency with accelerating voltage, it is difficult to achieve linear amplitude modulation with klystrons. Frequency modulation may, however, be readily accomplished by introducing a small modulating voltage in the cathode circuit.

10-23. The Haeff U-H-F Tube.- Figure 10-35 shows the essential structure and the circuit diagram of the Haeff ultra-high-frequency tube.[39] Like the klystron, this tube makes use of a cavity resonator to take energy from the electron beam, but the flow of electrons is controlled by means of a grid, instead of a resonator, and requires appreciable driving power. The circuit of Fig. 10-35 shows the use of the tube as an amplifier, but if the grid is excited by output taken from the cavity resonator, sustained oscillations may be produced. At 450 megacycles, Haeff reported an output of 110 watts at an efficiency of 35 per cent with an accelerating voltage of 6000 and a grid driving power of approximately 10 watts.

10-24. Hahn-Metcalf Velocity-modulated Tube.- Hahn and Metcalf have described velocity-modulated tubes in which velocity modulation of the electron beam is accomplished by means of grids.[40] In these tubes the velocity-modulation grid is prevented from affecting the number of electrons leaving the cathode by shielding the cathode from the grid. This may be

39. Haeff, A.V., Electronics, February 1939; A.V. Haeff and L.S. Nergaard, Proc. I.R.E., 28, 126 March (1940).
40. Hahn, W.C., and Metcalf, G.F., Proc. I.R.E., 27, 106 Feb. (1939).

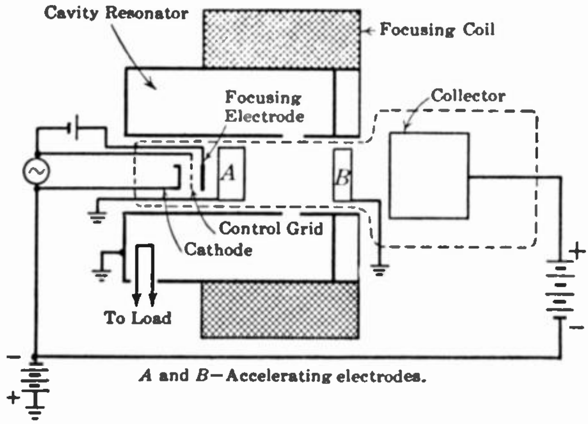


Fig. 10-35. Haeff U-H-F tube.

accomplished by using between the cathode and the velocity-modulation grid an additional grid, the potential of which is fixed relative to the cathode. In experimental tubes described by Hahn and Metcalf the velocity-modulation grid takes the form of a small cylinder placed between shielding baffles, as shown in Fig. 10-36. Energy is taken from the bunched electron beam by means of an anode similar in form to the grid, having a direct component of voltage of 10 to 30 volts. The electrodes numbered 5, 7, 9, and 11 in Fig. 10-36 are focusing electrodes that prevent spreading of the electron beam. They are maintained at a positive potential of about 30 volts relative to the cathode and operate in conjunction with the elements 2, 4, 6, 8, 10, 12, and 14, which are at a positive potential of 300 volts. The direct operating voltage of the grid is from 10 to 30 volts positive, and the voltage of the accelerating electrode 1 is about 60 volts. This tube is designed for use as an amplifier at frequencies ranging from 50 to 200 megacycles, but Hahn and Metcalf discuss and describe similar tubes used as oscillators at wavelengths down to 4.8 cm.

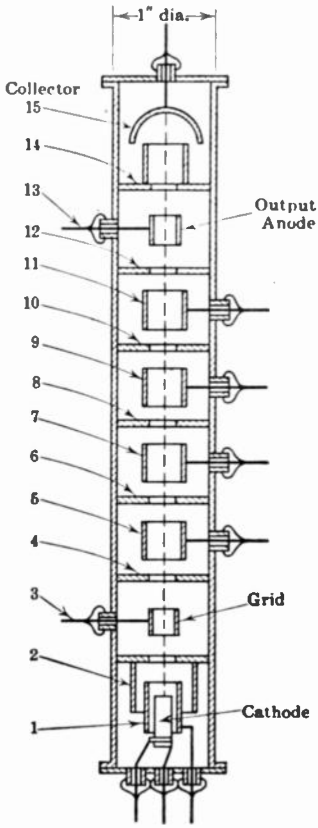


FIG. 10-36. Hahn and Metcalf Velocity-modulation Amplifier Tube.

BIBLIOGRAPHY

Extensive bibliographies will be found in the following articles:

"Vacuum Tubes as High-Frequency Oscillators," M.J.Kelly and A.L. Samuel, Bell System Tech. Jour., 14, 97 January (1935).

"Vacuum Tubes as High-Frequency Oscillators," E.D. McArthur and E.E. Spitzer, Proc. I.R.E., 19, 1971 November (1931).

"The Magnetron as a High-Frequency Generator," G.R. Kilgore, J.A.P., 8, 666 October (1937).

"The Operation of Electron Tubes at High Frequencies," H. Rothe, Proc. I.R.E., 28, 325 July 1940.

"The Characteristics of Negative-resistance Magnetrons," Hsu Chang and E.L. Chaffee, Proc. I.R.E., 28, 519 November (1940).

A collection of articles on ultra-high-frequency tubes and circuits is contained in "Radio at Ultra-High-Frequencies," RCA Institute Technical Press, 75 Varick Street, New York, New York.

Chapter II

TRANSMISSION LINES

Transmission lines are usually used as means for conveying power from a source to a load. In many ultra-high-frequency systems, however, there are other important uses; for example, as circuit elements in place of ordinary capacitors and inductors, for impedance matching, and for measuring wavelength. Furthermore, antennas may be treated as open-ended transmission lines, and the current distribution on them is very nearly that determined from transmission-line theory.

II-1. Fundamental Equations: Steady State.[1] - The conventional transmission-line equations are based upon the assumption of a uniform distribution of series resistance of R ohms and a series inductance of L henrys per unit loop[2] length, and a shunt conductance of G mhos and a shunt capacitance of C farads between "go" and "return" conductors, per unit length. Let

$\underline{Z} = R + j\omega L$, be the complex series impedance per unit loop length;

$\underline{Y} = G + j\omega C$, be the complex shunt admittance per unit loop length;

$\underline{Z}_o = R_o + jX_o$ be the complex impedance across line terminals at output end.

Here ω is the angular frequency $2\pi f$ of a <u>sinusoidal</u>[1] voltage $\underline{V}_1$ which is applied to the input (sending) end of the line.

The distance along the line is expressed in terms of the

---

1. It is essential to remember that the development is limited to steady-state conditions. This condition is introduced from the start by the use of the complex quantities $\underline{V}$, $\underline{I}$, $\underline{Z}_o$, etc. If transients, or wave forms not readily resolved into one or the sum of more than one sine wave, are to be studied, then the instantaneous currents and voltages, and R, L, G, C rather than $\underline{Z}$, $\underline{Y}$, etc. must be used. This discussion is continued in footnote 3.
2. By "loop" is meant the two wires of a two-wire line, the two conductors of a coaxial line, the one wire of a two-conductor system which uses the ground as the second conductor (assumed a perfect conductor), etc.

   A uniform line is assumed, and in the event the line is not uniform, substantial uniformity may often be achieved by transposing conductors at intervals or by other methods.

   Note particularly that $\underline{Z}$ and $\underline{Y}$ are not directly related; in particular $\underline{Y}$ is <u>not</u> the reciprocal of $\underline{Z}$.

variable x (measured in any convenient unit of length), whose origin is taken at the output (receiving) end. At any point x, the current and voltage are $\underline{I}$ and $\underline{V}$ respectively. The line is terminated in a load impedance, $\underline{Z}_o$.

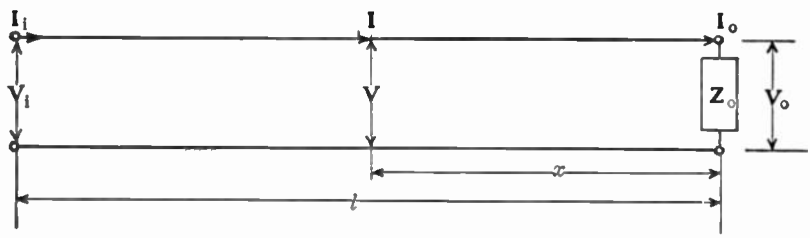


FIG.11-1. Transmission line with output (load) impedance Z_o.

It is clear that the vector drop in voltage per unit length along the line is equal to the $\underline{ZI}$ drop per unit length; and that the vector change in current per unit length along the line is $\underline{YV}$ per unit length. In equation form,

$$\frac{d\underline{V}}{dx} = \underline{Z}\,\underline{I} \tag{11-1}$$

and

$$\frac{d\underline{I}}{dx} = \underline{Y}\,\underline{V}: \tag{11-2}$$

By elimination of $\underline{I}$ and $\underline{V}$ in turn between the simultaneous differential equations, there result:

$$\frac{d^2\underline{V}}{dx^2} = \underline{ZY}\,\underline{V} = \underline{\gamma}^2\underline{V}, \tag{11-3}$$

$$\frac{d^2\underline{I}}{dx^2} = \underline{ZY}\,\underline{I} = \underline{\gamma}^2\underline{I}, \tag{11-4}$$

where $\underline{\gamma} \equiv \sqrt{\underline{ZY}}$ is the propagation constant. These second-order differential equations each have two independent solutions,[3]

3. If instantaneous voltage v and current i, rather than the complex equivalents $\underline{V}$ and $\underline{I}$ of sinusoidally varying quantities, had been considered, the statement that change in the voltage v per unit length at any point x on the line is equal to ri + Ldi/dt at that point would be written

$$\frac{\partial v}{\partial x} = ri + L\frac{\partial i}{\partial t} \tag{a}$$

which corresponds to (11-1), and the change in the current i per unit length of line at any point x equals Gv + C dv/dt at that point would be written which may be written either as $e^{\underline{\gamma} x}$ and $e^{-\underline{\gamma} x}$ or as $\cosh \underline{\gamma} x$ and $\sinh \underline{\gamma} x$ (hyperbolic cosine and hyperbolic sine of $\underline{\gamma} x$).

The following identities are useful in subsequent work:

$$\cosh \underline{\gamma} x = \frac{e^{\underline{\gamma} x} + e^{-\underline{\gamma} x}}{2} = 1 + \frac{\underline{\gamma}^2 x^2}{2!} + \frac{\underline{\gamma}^4 x^4}{4!} + \ldots\ldots;$$

$$\sinh \underline{\gamma} x = \frac{e^{\underline{\gamma} x} - e^{\underline{\gamma} x}}{2} = \underline{\gamma} x + \frac{\underline{\gamma}^3 x^3}{3!} + \frac{\underline{\gamma}^5 x^5}{5!} + \ldots\ldots;$$

$$e^{\underline{\gamma} x} = \cosh \underline{\gamma} x + \sinh \underline{\gamma} x = 1 + \underline{\gamma} x + \frac{\underline{\gamma}^2 x^2}{2!} + \frac{\underline{\gamma}^3 x^3}{3!} + \ldots\ldots;$$

$$e^{-\underline{\gamma} x} = \cosh \underline{\gamma} x - \sinh \underline{\gamma} x = 1 - \underline{\gamma} x + \frac{\underline{\gamma}^2 x^2}{2!} - \frac{\underline{\gamma}^3 x^3}{3!} + \ldots\ldots;$$

Instead of using differential equations to arrive at the functions representing voltage and current distribution, we could write

$$\underline{V} = \underline{V}_o + \int_o^x \underline{Z}\,\underline{I}\, dx\ ; \tag{11-5}$$

$$\underline{I} = \frac{\underline{V}_o}{\underline{Z}_o} + \int_o^x \underline{Y}\,\underline{V}\, dx\ ; \tag{11-6}$$

a pair of simultaneous integral equations stating respectively that the voltage $\underline{V}$ at point $\underline{x}$ is the sum of the output (receiver) voltage $\underline{V}_o$ plus the integrated $\underline{Z}\ \underline{I}$ drop between x and the output end; and that the current $\underline{I}$ at point x is the sum of the receiver current $\underline{V}_o/\underline{Z}_o$ plus the integrated shunt admittance current $\underline{Y}\ \underline{V}$ between x and the output end. By successive iteration:

Footnote Continued

$$\frac{\partial i}{\partial x} = Gv + C\frac{\partial v}{\partial t} \tag{b}$$

which is the more general form of (11-2).

Solving for i by differentiating (a) with respect to x and substituting from (b),

$$\frac{\partial^2 v}{\partial x^2} = RGv + (RC + LG)\frac{\partial v}{\partial t} + LC\frac{\partial^2 v}{\partial t^2} \tag{c}$$

There is an exactly similar equation for i. Now if v is assumed to vary sinusoidally with the time, then v can be considered the real part of $\underline{V}e^{j\omega t}$, which when substituted into (c) yields (11-3). But if v does not vary sinusoidally with t, as in the case of transients, pulses, etc., then a solution of (c) and of the corresponding equation for i must be obtained which satisfies all the conditions of the problem at hand. The reader will thus appreciate that the solutions for the steady alternating state obtained in most of this chapter do not necessarily apply to transients, pulse transmission, etc. A further discussion is contained in Sec. 11-12.

first inserting $\underline{V}_o$ as an approximation to $\underline{V}$ in (11-6); second, substituting the resulting value of $\underline{I}$ in (11-5) and solving for $\underline{V}$; third, substituting this value of $\underline{V}$ for the $\underline{V}$ in (11-6), and so on, there are obtained the two convergent series for $\underline{V}$ and $\underline{I}$:

$$\underline{V} = \underline{V}_o\left[\left(1 + \frac{\gamma^2 x^2}{2!} + \frac{\gamma^4 x^4}{4!} + \ldots\right) + \frac{\underline{Z}_K}{\underline{Z}_o}\left(\underline{\gamma} x + \frac{\gamma^3 x^3}{3!} + \frac{\gamma^5 x^5}{5!} + \ldots\right)\right] \qquad (11\text{-}7)$$

and

$$\underline{I} = \frac{\underline{V}_o}{\underline{Z}_o}\left[\left(1 + \frac{\gamma^2 x^2}{2!} + \frac{\gamma^4 x^4}{4!} + \ldots\right) + \frac{\underline{Z}_o}{\underline{Z}_K}\left(\underline{\gamma} x + \frac{\gamma^3 x^3}{3!} + \frac{\gamma^5 x^5}{5!} + \ldots\right)\right]. \qquad (11\text{-}8)$$

$\underline{Z}_K$ is defined as $\sqrt{\underline{Z}/\underline{Y}}$, and is called the characteristic impedance of the line. It is equal to the input impedance of a line of infinite length. It may be observed that the infinite series in (11-7) and (11-8) are equal to cosh $\underline{\gamma}x$ and sinh $\underline{\gamma}x$. It may be observed also that if $\underline{Z}_o$ is made equal to $\underline{Z}_K$, the two series in each of these two equations will have equal coefficients, and so may be combined, giving for their sum $e^{\underline{\gamma x}}$. Even when $\underline{Z}_o \neq \underline{Z}_K$, the series in (11-7) and (11-8) may be rearranged into $e^{\underline{\gamma x}}$ and $e^{-\underline{\gamma} x}$ groupings, the coefficient of the $e^{-\underline{\gamma x}}$ term being directly proportional to the mismatch between $\underline{Z}_o$ and $\underline{Z}_K$. The expressions are:[4]

$$\underline{V} = \underline{V}_o\left[\left(\frac{\underline{Z}_o + \underline{Z}_K}{2\underline{Z}_o}\right) e^{\underline{\gamma} x} + \left(\frac{\underline{Z}_o - \underline{Z}_K}{2\underline{Z}_o}\right) e^{-\underline{\gamma x}}\right] \qquad (11\text{-}9)$$

and

$$\underline{I} = \frac{\underline{V}_o}{\underline{Z}_o}\left[\left(\frac{\underline{Z}_K + \underline{Z}_o}{2\underline{Z}_K}\right) e^{\underline{\gamma x}} + \left(\frac{\underline{Z}_K - \underline{Z}_o}{2\underline{Z}_K}\right) e^{-\underline{\gamma x}}\right]. \qquad (11\text{-}10)$$

The alternative hyperbolic solutions are, from (11-7) and (11-8)

$$\underline{V} = \underline{V}_o\left[\cosh \underline{\gamma} x + \frac{\underline{Z}_K}{\underline{Z}_o} \sinh \underline{\gamma} x\right] \qquad (11\text{-}11)$$

and

$$\underline{I} = \frac{\underline{V}_o}{\underline{Z}_o}\left[\cosh \underline{\gamma} x + \frac{\underline{Z}_o}{\underline{Z}_K} \sinh \underline{\gamma} x\right] \qquad (11\text{-}12)$$

There are several reasons, in the design of communication circuits, for matching $\underline{Z}_o$ with Z_K, as follows:

1. The power transfer capacity of the line is usually at or near its maximum when $\underline{Z}_o = \underline{Z}_K$.

2. The line efficiency is usually near maximum when $\underline{Z}_o = \underline{Z}_K$.

4. It should be noted that it is not difficult to calculate a quantity such as $e^{\underline{\gamma x}}$ where $\underline{\gamma}$ is complex. If $\underline{\gamma} = \alpha + j\beta$ (α and β real) has been calculated and x is given, the αx and βx are known and

$$e^{\underline{\gamma x}} = e^{\alpha x} e^{j\beta x} = e^{\alpha x} \underline{/\beta x}$$

that is, $e^{\alpha x}$ is the modulus or magnitude of the complex quantity $e^{\underline{\gamma x}}$, and βx is the phase angle

3. Transient, echo and reflection effects are minimized when $\underline{Z}_o = \underline{Z}_K$.

In (11-9) and (11-10) the term $e^{\underline{\gamma}x}$ may be considered as the direct wave, attenuating in the direction of the load, and the term involving $e^{-\underline{\gamma}x}$ as the reflected wave, attenuating as it progresses away from the load. When the load impedance, $\underline{Z}_o$, is equal to $\underline{Z}_K$, equations (11-9) and (11-10) are simplified by the vanishing of their last terms, and for this condition they possess greater simplicity than the hyperbolic forms (11-11) and (11-12). For other values of Z_o, the forms of equations (11-11) and (11-12) are simpler, particularly if $\underline{Z}_o$ is equal to 0 or ∞.

All the foregoing equations have been expressed in terms of the variable distance x from the receiving end, but by the simple substitution of line length, ℓ, for x, the expressions for input voltage, $\underline{V}_1$, and current, $\underline{I}_1$, are obtained.

11-2. Formulas for Reference.- The following collection of the various forms of the transmission-line equations give explicit expressions for the various voltages and currents in terms of each of the others. These expressions follow readily from those derived above and are given in Table 11-I.

11-3. Approximate Expressions for γ and Z_K when $\omega L >> R$ and $\omega C >> G$.- We have defined the propagation constant $\underline{\gamma}$ and the characteristic impedance $\underline{Z}_K$ as:

$$\underline{\gamma} = \sqrt{\underline{Z}\,\underline{Y}} = \sqrt{(R + j\omega L)(G + j\omega C)}$$

$$= \sqrt{R\,G - \omega^2 L\,C + j\omega(L\,G + R\,C)}\;; \qquad (11\text{-}13)$$

$$\underline{Z}_K = \sqrt{\frac{\underline{Z}}{\underline{Y}}} = R_K + jX_K = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$

$$= \sqrt{\frac{R\,G + \omega^2 L\,C + j\omega(L\,G - R\,C)}{G^2 + \omega^2 C^2}} \qquad (11\text{-}14)$$

The propagation constant $\underline{\gamma}$ is in general complex, and its real and imaginary portions are denoted as

$$\underline{\gamma} = \alpha + j\beta. \qquad (11\text{-}15)$$

where α is called the <u>attenuation constant</u> and β the <u>phase-shift constant</u> of the line. Both α and β may be expressed as explicit functions of R, L, G, and C, but the expressions are long and cumbersome, and their use is not recommended. They are

$$\alpha = \left\{\frac{1}{2}[(RG - \omega^2 LC) + \sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)}]\right\}^{1/2}; \qquad (11\text{-}16)$$

$$\beta = \left\{\frac{1}{2}[-(RG - \omega^2 LC) + \sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)}]\right\}^{1/2} \qquad (11\text{-}17)$$

TABLE 11 - I

Currents, Voltages, and Impedances: Two-Conductor Transmission Line
(length = l, distance from output end to a point on the line = x)

QUANTITY	SYMBOL	VALUE [GENERAL CASE: OUTPUT(LOAD) IMPEDANCE = Z_0]	VALUE WHEN $Z_0 = Z_K$	VALUE WHEN $Z_0 = \infty$	VALUE WHEN $Z_0 = 0$
Output current	I_0	V_0/Z_0	$V_0/Z_K = I_1 e^{-\gamma l}$	0	$\frac{V_1}{Z_K} \sinh \gamma l = I_1/\cosh \gamma l$
Output voltage	V_0	$Z_0 I_0$	$Z_K I_0 = V_1 e^{-\gamma l}$	$V_1/\cosh \gamma l = Z_K I_1/\sinh \gamma l$	0
Load impedance (= impedance connected to output terminals)	Z_0	V_0/I_0	Z_K	∞	0
Input current	I_1	$I_0\left[\left(\frac{Z_0+Z_K}{2Z_K}\right)e^{\gamma l} - \left(\frac{Z_0-Z_K}{2Z_K}\right)e^{-\gamma l}\right]$ $= I_0\left(\cosh \gamma l + \frac{Z_0}{Z_K}\sinh \gamma l\right)$	$I_0 e^{\gamma l}$ $= I_0(\cosh \gamma l + \sinh \gamma l)$	$\frac{V_0}{Z_K}\sinh \gamma l = \frac{V_1}{Z_K}\tanh \gamma l$	$I_0 \cosh \gamma l = \frac{V_1}{Z_K \tanh \gamma l}$
Input voltage	V_1	$V_0\left[\left(\frac{Z_0+Z_K}{2Z_0}\right)e^{\gamma l} + \left(\frac{Z_0-Z_K}{2Z_0}\right)e^{-\gamma l}\right]$ $= V_0\left(\cosh \gamma l + \frac{Z_K}{Z_0}\sinh \gamma l\right)$	$V_0 e^{\gamma l}$ $= V_0(\cosh \gamma l + \sinh \gamma l)$	$V_0 \cosh \gamma l$	$Z_K I_0 \sinh \gamma l = Z_K I_1 \tanh \gamma l$
Input impedance	Z_1 $= V_1/I_1$	$Z_K\frac{(Z_0+Z_K)e^{\gamma l} + (Z_0-Z_K)e^{-\gamma l}}{(Z_0+Z_K)e^{\gamma l} - (Z_0-Z_K)e^{-\gamma l}}$ $= Z_K\frac{Z_0 + Z_K \tanh \gamma l}{Z_K + Z_0 \tanh \gamma l}$	Z_K	$Z_K/\tanh \gamma l$	$Z_K \tanh \gamma l$
Current at distance x from output end	I	(a) Substitute x for l in I_1 above or (b) use $I_1\left[\frac{Z_K \cosh \gamma x + Z_0 \sinh \gamma x}{Z_K \cosh \gamma l + Z_0 \sinh \gamma l}\right]$	$I_0 e^{\gamma x}$ $= I_1 e^{-\gamma(l-x)}$	$\frac{V_0}{Z_K}\sinh \gamma x$ $= I_1\frac{\sinh \gamma x}{\sinh \gamma l} = \frac{V_1}{Z_K}\frac{\sinh \gamma x}{\cosh \gamma l}$	$I_0 \cosh \gamma x$ $= I_1\frac{\cosh \gamma x}{\cosh \gamma l} = \frac{V_1}{Z_K}\frac{\cosh \gamma x}{\sinh \gamma l}$
Voltage at distance x from output end	V	(a) Substitute x for l in V_1 above, or (b) use $V_1\left[\frac{Z_0 \cosh \gamma x + Z_K \sinh \gamma x}{Z_0 \cosh \gamma l + Z_K \sinh \gamma l}\right]$	$V_0 e^{\gamma x}$ $= V_1 e^{-\gamma(l-x)}$	$V_0 \cosh \gamma x = V_1 \frac{\cosh \gamma x}{\cosh \gamma l}$	$Z_K I_0 \sinh \gamma x$ $= V_1\frac{\sinh \gamma x}{\sinh \gamma l} = Z_K I_1 \frac{\sinh \gamma x}{\cosh \gamma l}$
Impedance at distance x from output end looking toward output	V/I	Substitute x for l in Z_1	Z_K	$\frac{Z_K}{\tanh \gamma x}$	$Z_K \tanh \gamma x$

Numerical values for α and β are computed much more easily from the first of (11-13), if all four line parameters are to be used, and no approximations made. At very high frequencies ωL and ωC are so much larger than R and G respectively, that these loss terms may be dropped out in many (though not all) places with consequent simplification of the expressions for α, β and $\underline{Z}_K$. The approximate values of α and β so obtained are:

$$\alpha \doteq \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}} \qquad (11\text{-}18)$$

and

$$\beta \doteq \omega\sqrt{LC}\ . \qquad (11\text{-}19)$$

In many lines the last term in (11-18), involving G, is negligible in comparison with the term involving R, and may be dropped from the equation. Equation (11-18) should not be interpreted as stating that α is independent of the frequency, since R often varies with f.

Under the same conditions, we may write for the characteristic impedance

$$\underline{Z}_K \doteq \sqrt{\frac{L}{C}}\left[1 + \frac{1}{8}\left(\frac{R}{\omega L} - \frac{G}{\omega C}\right)^2 + \ldots\right] \underline{/ \frac{1}{2}\left(\frac{G}{\omega C} - \frac{R}{\omega L}\right)} \text{ complex ohms,} \qquad (11\text{-}19a)$$

and the angle is very small and may be considered zero for many lines.[5] The angle is given in radians. From this it is seen that as ω approaches ∞,

$$R_K \rightarrow \sqrt{\frac{L}{C}}, \qquad (11\text{-}20)$$

$$X_K \rightarrow 0. \qquad (11\text{-}21)$$

In Table 11-II are summarized the line properties. The case in which R/L = G/C has been listed because of its importance in those wire circuits to which inductance is added to approximate the condition.

11-4. **Lossless Line.**- Calculations based on a hypothetical lossless line are simpler than those where losses are taken into account even approximately, and often yield solutions which are quite adequate to the purpose at hand. At ultra-high frequencies, a line may be very short physically yet electrically long (e.g., several wavelengths) and have losses which are but an insignificant fraction of the power transmitted. For the lossless condition, where R = 0 and G = 0, the propagation constant and the characteristic impedance obviously reduce to

$$\underline{\gamma} = \alpha + j\beta = 0 + j\omega\sqrt{LC} \qquad (11\text{-}22)$$

and

$$\underline{Z}_K = R_K + jX_K = \sqrt{\frac{L}{C}} + j0. \qquad (11\text{-}23)$$

In Table 11-III have been collected some of the conditions under which a lossless line acts like a circuit element. All the results have been obtained by first substituting $\gamma = j\beta$ in formulas given in Table 11-I.

11-5. **Methods of Computation.**- From the value given in Table 11-I for the voltage along a transmission line terminated by a load whose impedance is equal to $\underline{Z}_K$, it is seen that the

5. An interesting case in which the angle may not be considered zero at high frequencies is given by L. S. Nergaard and Bernard Salzberg, "Resonant Impedance of Transmission Lines," Proc. I.R.E., Vol. 27, No. 9, Sept. 1939.

TABLE 11-II

Uniform Two-Conductor Transmission Line - Properties of the Line

[R and L are resistance and inductance respectively per unit length (two conductors); G and C are conductance and capacitance respectively per unit length of line (between two conductors)]

Quantity	Symbol	Value (General)	Approximate value: special case $\omega L >> R$ and $\omega C >> G$	Value: Special case $\frac{R}{L} = \frac{G}{C}$
Line impedance per unit length (two conductors)	$\underline{Z} = Z \angle\theta_Z$	$= R + j\omega L = \sqrt{R^2 + \omega^2 L^2} \angle\tan^{-}(\omega L/R)$		$\frac{L}{C}\underline{Y}$
Line admittance per unit length (between two conductors)	$\underline{Y} = Y \angle\theta_Y$	$= G + j\omega C = \sqrt{G^2 + \omega^2 C^2} \angle\tan^{-1}(\omega C/G)$		$\frac{C}{L}\underline{Z}$
Characteristic impedance	$\underline{Z_K} = Z_K \angle\theta_K$	$= \sqrt{\frac{\underline{Z}}{\underline{Y}}} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} = \sqrt[4]{\frac{R^2 + \omega^2 L^2}{G^2 + \omega^2 C^2}} \angle\frac{1}{2}(\theta_z - \theta_y)$	$\sqrt{\frac{L}{C}}$	$\sqrt{\frac{L}{C}}$
Propagation constant	$\underline{\gamma} = \alpha + j\beta$	$= \sqrt{\underline{Y}\,\underline{Z}} = \underline{Y}\,\underline{Z_K} = \frac{\underline{Z}}{\underline{Z_K}} = \sqrt{(R + j\omega L)(G + j\omega C)}$		$\underline{Z}\sqrt{\frac{G}{R}} = \underline{Y}\sqrt{\frac{R}{G}}$
Attenuation constant	α	$= \sqrt{YZ}\,\cos\frac{1}{2}(\theta_y + \theta_z) = \sqrt{YZ + \frac{1}{2}(RG - \omega^2 LC)}$	$\frac{1}{2}R\sqrt{\frac{C}{L}} + \frac{1}{2}G\sqrt{\frac{L}{C}} = \frac{1}{2}\frac{R}{Z_K}(1 + \frac{G}{R}\frac{L}{C})$	$\sqrt{RG}$
Phase-shift constant	β	$= \sqrt{YZ}\,\sin\frac{1}{2}(\theta_y + \theta_z) = \sqrt{YZ - \frac{1}{2}(RG - \omega^2 LC)} = \sqrt{(\omega^2 LC - RG) + \alpha^2}$	$\omega\sqrt{LC}$	$\omega\sqrt{LC}$
Phase velocity		$= \frac{\omega}{\beta}$	$\frac{1}{\sqrt{LC}}$	$\frac{1}{\sqrt{LC}}$

TABLE 11-III

Use of Transmission Line as a Circuit Element

Assumptions: R, G, α are effectively zero (lossless line), hence $\underline{\gamma} = j\beta = j\omega\sqrt{LC}$ $\underline{Z}_K = \sqrt{L/C}$ pure resistance

Output (load) impedance $\underline{Z}_o$	Length of line ℓ	Input impedance $\underline{Z}_1$	Equivalent circuit element
0	any	$j\underline{Z}_K \tan \beta\ell$	Inductor when $\beta\ell$ falls in first or third quadrants Capacitor when $\beta\ell$ falls in second or fourth quadrants.
∞	any	$-j\underline{Z}_K/\tan \beta\ell$	Capacitor when $\beta\ell$ falls in first or third quadrants Inductor when $\beta\ell$ falls in second or fourth quadrants.
Real	$(2N + 1)\frac{\lambda}{4}$ (N any integer)	$\left(\frac{\underline{Z}_K^2}{\underline{Z}_o^2}\right)\underline{Z}_o$	Transformer: $\underline{V}_1 \underline{I}_1 = \underline{V}_o \underline{I}_o$ and $\frac{\underline{V}_o}{\underline{V}_1} = \frac{\underline{I}_1}{\underline{I}_o} = \frac{\underline{Z}_o}{\underline{Z}_K}$ = equivalent turn ratio. Note that $\underline{Z}_1$ is square of the reciprocal of the turn ratio multiplied by $\underline{Z}_o$
$\underline{Z}_o$	same	same	Same, except that equivalent turn ratio is complex

variation is proportional to an exponential, i.e.,

$$\underline{V} = \underline{V}_1 e^{-\underline{\gamma}(\ell - x)} \tag{11-24}$$

A point at distance x from the output end is the same as one at distance ℓ-x from the input end. On a line of infinite length, the term ℓ-x would have a rather tenuous meaning, but the equation would still hold true if ℓ-x is replaced by the distance from the input end. We have defined $\underline{\gamma}$ as

$$\underline{\gamma} = \alpha + j\beta \tag{11-25}$$

and so eq. (11-24) may be written as

$$\begin{aligned} \underline{V} &= \underline{V}_1 \; [e^{-\alpha(\ell - x)} \; e^{-j\beta(\ell - x)}] \\ &= \underline{V}_1 \; e^{-\alpha(\ell - x)}[\cos \beta(\ell - x) - j \sin \beta(\ell - x)] \end{aligned} \tag{11-26}$$

Since both α and β are real, it is apparent that α is an attenuation constant and β a phase constant, at any fixed frequency. Both the exponential function and the trigonometric functions may of course be evaluated readily from slide rule or tables. The arguments are in radians.

When the line formula to be applied contains hyperbolic functions, for example

$$\underline{V}_1 = \underline{V}_o \cosh \underline{\gamma}\ell + \frac{\underline{Z}_K}{\underline{Z}_o} \sinh \underline{\gamma}\ell, \tag{11-27}$$

the complex hyperbolic functions are best evaluated by expansion in terms of real functions. The following identities are frequently useful:

$$\cosh (a + jb) = \cosh a \cos b + j \sinh a \sin b; \qquad (11\text{-}28)$$

$$\sinh (a + jb) = \sinh a \cos b + j \cosh a \sin b. \qquad (11\text{-}29)$$

These identities are easily proven by the substitution of the exponential forms for all the functions. Values of the real hyperbolic functions are given in Table 11-IV, for arguments lying between 0 and 3.00. For larger arguments the cosh and sinh functions are approximately equal to each other and to half the positive exponential of the same argument.

If the complex argument $\underline{\gamma}\ell$ has a magnitude or absolute value less than about unity, then the direct use of the appropriate series expression for the function may be convenient, but for large arguments the series will be found to converge so slowly that a great many terms are required to give accurate results. The use of the series with a complex argument of course will make each term complex, as a general thing.

Kennelly's tables or charts of complex hyperbolic functions may be referred to if available.[6]

11-6. Example. Voltage and Current Distribution.- Let it be required to determine and plot the complex voltage and current distribution on a transmission line 8 miles long with open end ($\underline{Z}_o = \infty$), the line parameters per mile being

$$\begin{aligned} R &= 108.8 \text{ ohms per mile,} \\ \omega L &= 335.6 \text{ ohms per mile,} \\ G &= 0, \\ \omega C &= 2.024 \times 10^{-3} \text{ mho per mile.} \end{aligned}$$

The open-end voltage is one volt.

Solution. The propagation constant $\underline{\gamma}$ and the characteristic impedance $\underline{Z}_K$ are first calculated.

$$\underline{\gamma} = \sqrt{\underline{ZY}} = \sqrt{(108.8 + j335.6)(0 + j0.002024)}$$

$$= 0.1320 + j0.8348 \text{ complex numeric per mile.}$$

$$\underline{Z}_K = \sqrt{\frac{\underline{Z}}{\underline{Y}}} = \sqrt{\frac{108.8 + j335.6}{0 + j0.002024}}$$

$$= 417.5 \underline{/-7.98^\circ} \text{ complex ohms.}$$

Equations for $\underline{V}$ and $\underline{I}$ from Table 11-I will be used, and the functions evaluated by expansion in real functions.

6. "Tables of Complex Hyperbolic and Circular Functions," A.E. Kennelly, Cambridge, 1921. "Chart Atlas of Complex Hyperbolic and Circular Functions," A.E. Kennelly, Cambridge, 1924.

TABLE 11-IV

Hyperbolic Sines and Cosines

Argument	Function	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	sinh	0.0000	0.0100	0.0200	0.0300	0.0400	0.0500	0.0600	0.0701	0.0801	0.0901
	cosh	1.0000	1.0001	1.0002	1.0005	1.0008	1.0013	1.0018	1.0025	1.0032	1.0041
0.1	sinh	0.1002	0.1102	0.1203	0.1304	0.1405	0.1506	0.1607	0.1708	0.1810	0.1911
	cosh	1.0050	1.0061	1.0072	1.0085	1.0098	1.0113	1.0128	1.0145	1.0162	1.0181
0.2	sinh	0.2013	0.2115	0.2218	0.2320	0.2423	0.2526	0.2629	0.2733	0.2837	0.2941
	cosh	1.0201	1.0221	1.0243	1.0266	1.0289	1.0314	1.0340	1.0367	1.0395	1.0423
0.3	sinh	0.3045	0.3150	0.3255	0.3360	0.3466	0.3572	0.3678	0.3785	0.3892	0.4000
	cosh	1.0453	1.0484	1.0516	1.0549	1.0584	1.0619	1.0655	1.0692	1.0731	1.0770
0.4	sinh	0.4108	0.4216	0.4325	0.4434	0.4543	0.4653	0.4764	0.4875	0.4986	0.5098
	cosh	1.0811	1.0852	1.0895	1.0939	1.0984	1.1030	1.1077	1.1125	1.1174	1.1225
0.5	sinh	0.5211	0.5324	0.5438	0.5552	0.5666	0.5782	0.5897	0.6014	0.6131	0.6248
	cosh	1.1276	1.1329	1.1383	1.1438	1.1494	1.1551	1.1609	1.1669	1.1730	1.1792
0.6	sinh	0.6367	0.6485	0.6605	0.6725	0.6846	0.6967	0.7090	0.7213	0.7336	0.7461
	cosh	1.1855	1.1919	1.1984	1.2051	1.2119	1.2188	1.2258	1.2330	1.2402	1.2476
0.7	sinh	0.7586	0.7712	0.7838	0.7966	0.8094	0.8223	0.8353	0.8484	0.8615	0.8748
	cosh	1.2552	1.2628	1.2706	1.2785	1.2865	1.2947	1.3030	1.3114	1.3199	1.3286
0.8	sinh	0.8881	0.9015	0.9150	0.9286	0.9423	0.9561	0.9700	0.9840	0.9981	1.0122
	cosh	1.2374	1.3464	1.3555	1.3647	1.3740	1.3835	1.3932	1.4029	1.4128	1.4229
0.9	sinh	1.0265	1.0409	1.0554	1.0700	1.0847	1.0995	1.1144	1.1294	1.1446	1.1598
	cosh	1.4331	1.4434	1.4539	1.4645	1.4753	1.4862	1.4973	1.5085	1.5199	1.5314
1.0	sinh	1.1752	1.1907	1.2063	1.2220	1.2379	1.2539	1.2700	1.2862	1.3025	1.3190
	cosh	1.5431	1.5549	1.5669	1.5790	1.5913	1.6038	1.6164	1.6292	1.6421	1.6552
1.1	sinh	1.3356	1.3524	1.3693	1.3863	1.4035	1.4208	1.4382	1.4558	1.4735	1.4914
	cosh	1.6685	1.6820	1.6956	1.7093	1.7233	1.7374	1.7517	1.7662	1.7808	1.7956
1.2	sinh	1.5095	1.5276	1.5460	1.5645	1.5831	1.6019	1.6209	1.6400	1.6593	1.6788
	cosh	1.8107	1.8258	1.8412	1.8568	1.8725	1.8884	1.9045	1.9208	1.9373	1.9540
1.3	sinh	1.6984	1.7182	1.7381	1.7583	1.7786	1.7991	1.8198	1.8406	1.8617	1.8829
	cosh	1.9709	1.9880	2.0053	2.0228	2.0404	2.0583	2.0764	2.0947	2.1132	2.1320
1.4	sinh	1.9043	1.9259	1.9477	1.9697	1.9919	2.0143	2.0369	2.0597	2.0827	2.1059
	cosh	2.1509	2.1700	2.1894	2.2090	2.2288	2.2488	2.2691	2.2896	2.3103	2.3312

Hyperbolic Sines and Cosines

Argument	Function	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	sinh	2.1293	2.1529	2.1768	2.2008	2.2251	2.2496	2.2743	2.2993	2.3245	2.3499
	cosh	2.3524	2.3738	2.3955	2.4174	2.4395	2.4619	2.4845	2.5074	2.5305	2.5538
1.6	sinh	2.3756	2.4015	2.4276	2.4540	2.4806	2.5075	2.5346	2.5620	2.5896	2.6175
	cosh	2.5775	2.6013	2.6255	2.6499	2.6746	2.6995	2.7247	2.7502	2.7760	2.8020
1.7	sinh	2.6456	2.6740	2.7027	2.7317	2.7609	2.7904	2.8202	2.8503	2.8806	2.9112
	cosh	2.8283	2.8549	2.8818	2.9090	2.9364	2.9642	2.9922	3.0206	3.0493	3.0782
1.8	sinh	2.9422	2.9734	3.0049	3.0367	3.0689	3.1013	3.1340	3.1671	3.2005	3.2341
	cosh	3.1075	3.1371	3.1669	3.1972	3.2277	3.2585	3.2897	3.3212	3.3530	3.3852
1.9	sinh	3.2682	3.3025	3.3372	3.3722	3.4075	3.4432	3.4792	3.5156	3.5523	3.5894
	cosh	3.4177	3.4506	3.4838	3.5173	3.5512	3.5855	3.6201	3.6551	3.6904	3.7261
2.0	sinh	3.6269	3.6647	3.7028	3.7414	3.7803	3.8196	3.8593	3.8993	3.9398	3.9806
	cosh	3.7622	3.7987	3.8355	3.8727	3.9103	3.9483	3.9867	4.0255	4.0647	4.1043
2.1	sinh	4.0219	4.0635	4.1056	4.1480	4.1909	4.2342	4.2779	4.3221	4.3666	4.4117
	cosh	4.1443	4.1847	4.2256	4.2668	4.3085	4.3507	4.3932	4.4362	4.4797	4.5236
2.2	sinh	4.4571	4.5030	4.5494	4.5962	4.6434	4.6912	4.7394	4.7880	4.8372	4.8868
	cosh	4.5679	4.6127	4.6580	4.7037	4.7499	4.7966	4.8437	4.8914	4.9395	4.9881
2.3	sinh	4.9370	4.9876	5.0387	5.0903	5.1425	5.1951	5.2483	5.3020	5.3562	5.4109
	cosh	5.0372	5.0868	5.1370	5.1876	5.2388	5.2905	5.3427	5.3954	5.4487	5.5026
2.4	sinh	5.4662	5.5221	5.5785	5.6354	5.6929	5.7510	5.8097	5.8689	5.9288	5.9892
	cosh	5.5569	5.6119	5.6674	5.7235	5.7801	5.8373	5.8951	5.9535	6.0125	6.0721
2.5	sinh	6.0502	6.1118	6.1741	6.2369	6.3004	6.3645	6.4293	6.4946	6.5607	6.6274
	cosh	6.1323	6.1931	6.2545	6.3166	6.3793	6.4426	6.5066	6.5712	6.6365	6.7024
2.6	sinh	6.6947	6.7628	6.8315	6.9009	6.9709	7.0417	7.1132	7.1854	7.2583	7.3319
	cosh	6.7690	6.8363	6.9043	6.9729	7.0423	7.1123	7.1831	7.2546	7.3268	7.3998
2.7	sinh	7.4063	7.4814	7.5572	7.6338	7.7112	7.7894	7.8683	7.9480	8.0285	8.1098
	cosh	7.4735	7.5479	7.6231	7.6991	7.7758	7.8533	7.9316	8.0106	8.0905	8.1712
2.8	sinh	8.1919	8.2749	8.3586	8.4432	8.5287	8.6150	8.7021	8.7902	8.8791	8.9689
	cosh	8.2527	8.3351	8.4182	8.5022	8.5871	8.6728	8.7594	8.8469	8.9352	9.0244
2.9	sinh	9.0596	9.1512	9.2437	9.3371	9.4315	9.5268	9.6231	9.7203	9.8185	9.9177
	cosh	9.1146	9.2056	9.2976	9.3905	9.4844	9.5792	9.6749	9.7716	9.8693	9.9680

TABLE 11-V

Computation of Voltage and Current Distribution along a Line; $\underline{Z}_0 = \infty$,
$\alpha = 0.1320$ per mile, $\beta = 0.8348$ per mile, $\underline{Z}_K = 417.5\underline{/-7.98^\circ}$ complex ohms,
$\ell = 8$ miles, $\underline{V}_0 = 1\underline{/0^\circ}$ volt

x (miles)	αx (numeric)	βx (radians)	cosh αx	cos βx	sinh αx	sin βx	$\underline{V} = \cosh \underline{\gamma}x$ (complex volts)	Symbol on Fig.11-2
0	0	0	1	1	0	0	1 + j0	$\underline{V}_0$
1	0.1320	0.8348	1.0087	0.6712	0.1324	0.7413	0.6770 + j0.09815	$\underline{V}_1$
2	0.2640	1.670	1.0341	-0.0890	0.2774	0.9951	-0.09203 + j0.2760	$\underline{V}_2$
3	0.3959	2.504	1.0794	-0.8034	0.4059	0.5952	-0.8672 + j0.2416	$\underline{V}_3$
4	0.5279	3.339	1.1424	-0.9806	0.5527	-0.1958	-1.1202 - j0.1082	$\underline{V}_4$
5	0.6599	4.174	1.2257	-0.5131	0.7089	-0.8583	-0.6289 - j0.6077	$\underline{V}_5$
6	0.7918	5.008	1.3302	0.2909	0.8772	-0.9567	0.3870 - j0.8392	$\underline{V}_6$
7	0.9238	5.844	1.4579	0.9049	1.0609	-0.4256	1.3193 - j0.4515	$\underline{V}_7$
8	1.0558	6.678	1.6111	0.9226	1.2632	0.3848	1.4864 + j0.4861	$\underline{V}_8$

x (miles)	$\sinh \underline{\gamma}x$	$\underline{I} = \frac{\sinh \underline{\gamma}x}{\underline{Z}_K}$ (complex milliamp)	Symbol on Fig. 11-2
0	0	0	
1	0.08887 + j0.7477	-3.79 + j 180.31	$\underline{I}_1$
2	-0.02469 + j1.0290	-40.08 + j 243.26	$\underline{I}_2$
3	-0.3261 + j0.6425	-98.72 + j 141.55	$\underline{I}_3$
4	-0.5420 - j0.2237	-121.12 - j 71.09	$\underline{I}_4$
5	-0.3633 - j1.052	- 51.18 - j 261.61	$\underline{I}_5$
6	0.2519 - j1.2727	102.09 - j 292.50	$\underline{I}_6$
7	0.9601 - j0.6204	248.38 - j 115.23	$\underline{I}_7$
8	1.1654 + j0.6214	245.76 + j 186.17	$\underline{I}_8$

Thus we have:

$$\underline{V} = \underline{V}_0 \cosh \underline{\gamma}x$$
$$= \underline{V}_0[(\cosh \alpha x \cos\beta x) + j(\sinh \alpha x \sin\beta x)]$$
$$= (\cosh 0.1320x \cos 0.8348x) + j(\sinh 0.1320x \sin 0.8348x) \text{ complex volts.}$$

$$\underline{I} = \frac{\underline{V}_0}{\underline{Z}_0} \sinh \underline{\gamma}x$$
$$= \frac{\underline{V}_0}{\underline{Z}_K} [(\sinh \alpha x \cos\beta x) + j(\cosh \alpha x \sin\beta x)]$$
$$= (0.002372 + j0.0003326) [(\sinh 0.1320x \cos 0.8348x) + j(\cosh 0.1320x \sin 0.8348x)] \text{ complex amperes,}$$

where $\underline{V}_0$ has been taken as equal to $1\underline{/0^\circ}$ volt. The numerical work is presented in tabular form in Table 11-V.

The complex voltage and current functions are plotted in Fig. 11-2. It is seen that the line is slightly more than one wavelength long, as is indicated by the voltage spiral making slightly more than one revolution about the origin for the 8 miles.

If there were no losses on the line, this voltage spiral would degenerate into a horizontal straight line joining

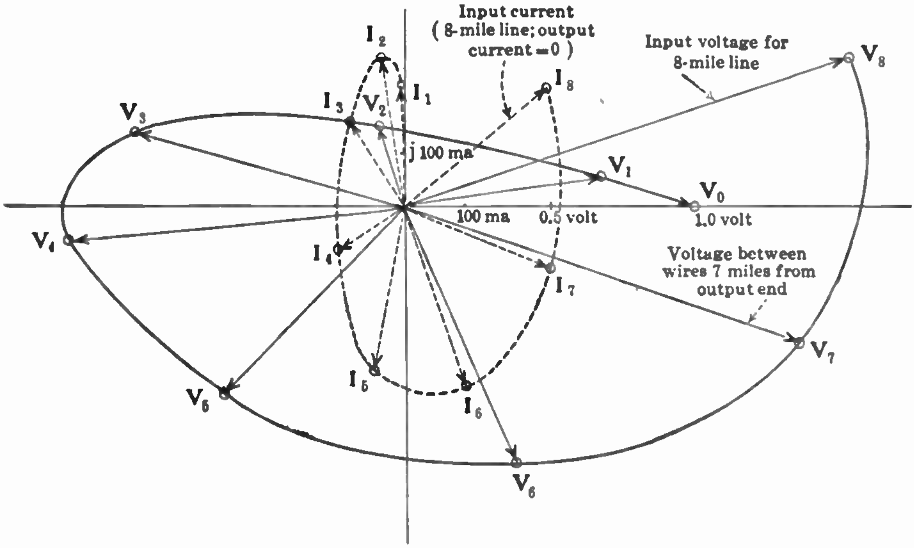


FIG. 11-2. Complex voltage and current distribution on an open-ended transmission line of length a little more than one wave-length. The attenuation of this line is relatively high ($\alpha = 0.1320$ per mile).

the points 1 + j0 and -1 + j0. The current spiral would degenerate into a vertical straight line.

Another way of showing graphically the distribution of current and voltage is illustrated in Fig. 11-3 and Fig. 11-4.

11-7. Direct (Incident) and Reflected Waves.- The distribution of voltage in the foregoing example was given by the equation

$$\underline{V} = \underline{V}_o \cosh (0.1320 + j0.8348)x$$

which may alternatively be expanded as

$$\underline{V} = \frac{\underline{V}_o}{2} (e^{0.1320x} e^{j0.8348x} + e^{-0.1320x} e^{-j0.8348x})$$

$$= \frac{\underline{V}_o}{2} [e^{0.1320x} (\cos 0.8348x + j \sin 0.8348x) + e^{-0.1320x} (\cos 0.8348x - j \sin 0.8348x)].$$

The two portions represent two waves; the first a sinusoid multiplied by an exponential increasing with x, and the second als a sinusoid, but multiplied by an exponential decreasing as x increases. Since any wave traveling along a line must undergo attenuation in the direction of its travel, the two portions must represent waves traveling in opposite directions; in other

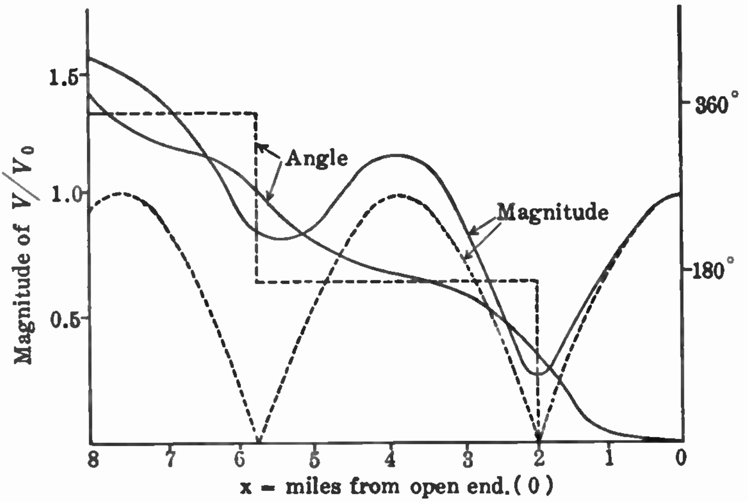


FIG. 11-3. Distribution of voltage magnitude and angle along an open line. The dotted curves are for corresponding lossless line.

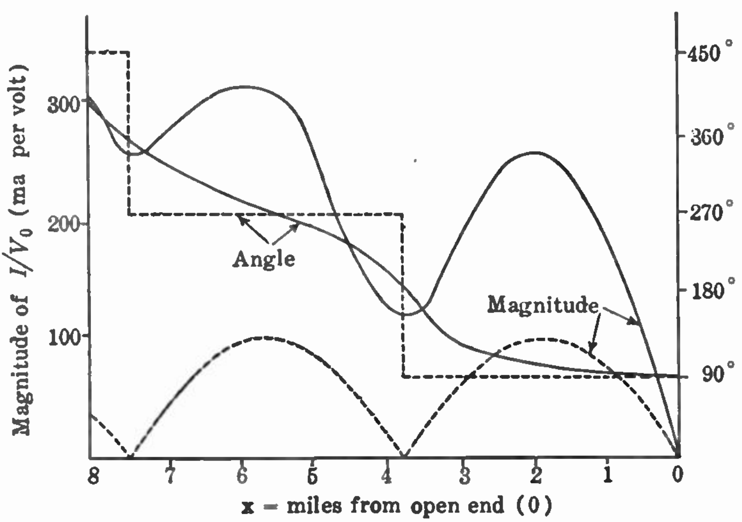


FIG. 11-4. Distribution of current magnitude and angle along an open line. The dotted curves are for corresponding lossless line.

words, a direct or incident and a reflected wave. Here ($\underline{Z}_o = \infty$) the reflection is complete, and the incident and reflected waves are equal at the open end. Farther back on the line, the direct wave is the larger. The summation of the two produces a standing wave pattern, which is easily recognized in the "magnitude" curves of Fig. 11-3. The two waves may be represented by two vectors rotating in opposite directions, with angle of rotation proportional to x, one increasing and the other shrinking.

The reflected wave may be reduced to zero by making the load impedance $\underline{Z}_o = \underline{Z}_K$. There will then be no standing wave pattern, and the curve of voltage magnitude will be a simple exponential curve, rising with increasing values of x, as indicated in Table 11-I.

Consider the current $\underline{I}$ at any point x; by Table 11-I

$$\underline{I} = \underline{A}e^{\gamma x} - \underline{B}e^{-\gamma x} \tag{11-30}$$

where $\underline{A}$ and $\underline{B}$ are the complex quantities, independent of x, which are indicated in the table. Write $\underline{A} = A\angle\theta_A$ and $\underline{B} = B\angle\theta_B$, and substitute $\underline{\gamma} = \alpha + j\beta$, then:

$$\underline{I} = Ae^{\alpha x}\angle\beta x + \theta_A - Be^{-\alpha x}\angle -\beta x + \theta_B \,. \tag{11-31}$$

Now, since a complex current or voltage is the equivalent of an actual instantaneous sine-wave current or voltage of the same magnitude as the complex quantity and with a phase angle equal to that of the complex quantity, the instantaneous current i corresponding to (11-31) can be written

$$i = Ae^{\alpha x}\cos(\omega t + \beta x + \theta_A) - Be^{-\alpha x}\cos(\omega t - \beta x + \theta_B). \tag{11-32}$$

When x is fixed, each term represents a sine wave of frequency $\omega/2\pi$, indicating that i is a sinusoidal current at x whatever x may be. But if t is held constant and x is varied--this is equivalent to the observer moving along the line with the current frozen at the distribution obtaining at the given instant--a cosine variation multiplied by $e^{\alpha x}$ or $e^{-\alpha x}$ will be observed. A repetition of this procedure an instant later will show that the A component appears to have moved toward the output end, whereas the B component appears to have moved in the opposite direction. Thus (11-30) is the complex equivalent of two waves, the direct or incident A wave (incident on $\underline{Z}_o$) and the reflected B wave (moving away from $\underline{Z}_o$). Figure 11-5 illustrates the distribution of the B wave with distance (x) for several values of t. Traveling waves of the type discussed here play an important part in many aspects of ultra-high-frequency work, and the reader not acquainted with them should plot the A term of (11-32) against x for several different values of t, until he feels that he has a good appreciation of the significance of

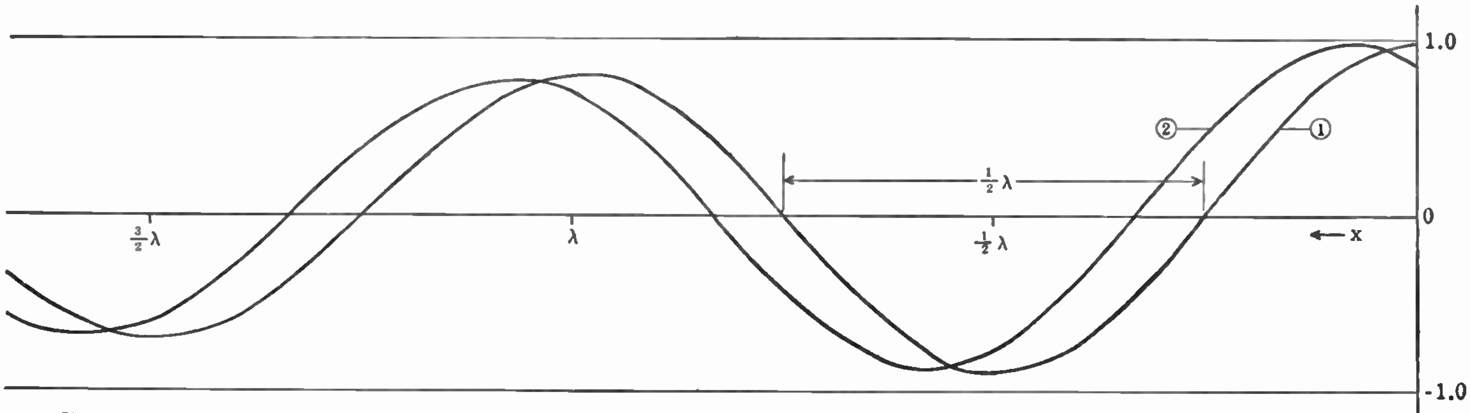


FIG. 11-5. Damped traveling wave: $e^{-\alpha x}\cos(\omega t - \beta x + \theta_E)$ plotted against x for two values of t — ① $t = 0$, ② $t = \frac{1}{12}$ T later. Wave travels from right to left with velocity λ/T.

At any fixed point there is a sinusoidal variation with time t; period T is change in t required to change ωt by 2π $(\omega = 2\pi f = 2\pi/T)$.

At any fixed time there is a damped sinusoidal variation with x; wave length λ is change in x required to change βx by 2π $(\beta = 2\pi/\lambda)$.

the expression; a similar plot of the B term (as in Fig. 11-5) should be made to show that it moves in the direction opposite to the first case.

The change in x required to change βx by 2π is a wavelength (λ); hence $\beta = 2\pi/\lambda$.

Consider a time t_1 and a point x_1, and a time t_2 and point x_2, related by $\omega t_1 + \beta x_1 = \omega t_2 + \beta x_2$. Then $\cos(\omega t_1 + \beta x_1 + \theta_A) = \cos(\omega t_2 + \beta x_2 + \theta_A)$, and states physically that the value of the cosine at time t_1 and point x_1 has moved to point x_2 at time t_2. The velocity of propagation is

$$\frac{x_1 - x_2}{t_2 - t_1} = \frac{\omega}{\beta} \qquad (11\text{-}33)$$

This is known as the velocity of phase propagation because it refers to the rate at which a given value of the cosine in a steady-state case appears to move; that is, the phase $\omega t_1 + \beta x_1 + \theta_A$ of the cosine at time t_1 and point x_1 is the same as the phase $\omega t_2 + \beta x_2 + \theta_A$ of the cosine at time t_2 and point x_2.

Velocity of phase propagation is indicated in Table 11-II under the brief table "phase velocity." It is desirable that phase velocity be independent of frequency, in order that waves of different frequencies maintain the same relative phases with respect to one another in steady-state transmission.[7]

11-8. Standing Waves.- The magnitude and phase angle of $\underline{I} = I/\theta_I$ in (11-31) can be obtained by combining the two complex quantities on the right. These are

$$I = \sqrt{A^2 e^{2\alpha x} + B^2 e^{-2\alpha x} - 2\,AB\cos(2\beta x + \theta_A - \theta_B)} \qquad (11\text{-}34)$$

and

$$\tan\theta_I = \frac{A\sin(\beta x + \theta_A) + Be^{-2\alpha x}\sin(\beta x - \theta_B)}{A\cos(\beta x + \theta_A) - Be^{-2\alpha x}\cos(\beta x - \theta_B)} \qquad (11\text{-}35)$$

For a particular case these have been plotted in Fig. 11-4, in which the solid "magnitude" curve is of the type of I above, and the solid "angle" curve is of the type of θ_I above.

A special case of much interest is when A = B and $\theta_A = \theta_B$ or $\theta_A = \theta_B + \pi$, and the line is lossless ($\alpha = 0$). It is seen from (11-34) that I becomes

$$I = \sqrt{A^2 + B^2 \pm 2\,AB\cos 2\beta x} \qquad (11\text{-}35a)$$

and I, the magnitude of the instantaneous i, will always be zero at certain points. Turning to the expression (11-32) for the

7. Velocity of phase propagation is not applicable to transients, etc. As its name implies, it indicates the speed with which a given phase of a sine wave appears to travel. It does not imply that energy is transmitted at this speed. Phase velocity may be greater than the velocity of light, as is the case in wave guides (Ch. 14).

instantaneous current i corresponding to the complex $\underline{I}$, and assuming A = B, and $\theta_A = \theta_B$, and $\alpha = 0$,

$$i = A[\cos(\omega t + \theta_A + \beta x) - \cos(\omega t + \theta_A - \beta x)] \tag{11-36}$$

$$= 2A \sin \beta x \sin(\omega t + \theta_A)$$

This is a standing wave, which results from combining an incident wave and a reflected wave of the same magnitude. At each point x the current i varies sinusoidally with the time. But the magnitude I of the current varies with x in such a manner that at some points (called nodes) I is always zero, and at some points I is always a maximum compared with the value of I at other points. Figure 11-6 illustrates (11-36).

In (11-30), $\underline{A}$ and $\underline{B}$ represent quantities given in Table 11-I. It is not necessary to introduce these quantities so long as it is known that they do not depend on x. The reader can show that if A = B, without any limitation on the phase angles of $\underline{A}$ and $\underline{B}$, standing waves will result.

Since Table 11-I shows that $\underline{A}$ and $\underline{B}$ actually stand for

$$\underline{A} = \underline{I}_o\left(\frac{\underline{Z}_o + \underline{Z}_K}{2\underline{Z}_o}\right) \tag{11-37}$$

and

$$\underline{B} = \underline{I}_o\left(\frac{\underline{Z}_o - \underline{Z}_K}{2\underline{Z}_o}\right) \tag{11-38}$$

it follows that if $\underline{Z}_K$ is real (as it is for most practical purposes at high frequencies--see Table 11-II), then any pure reactance jX_o for load ($\underline{Z}_o$) will result in standing waves. For A will equal B even though the phase angles of $\underline{A}$ and $\underline{B}$ may differ. The condition includes $\underline{Z}_o = 0$ and $\underline{Z}_o = \infty$.

11-9. Quarter-Wave Line as a Transformer. - It will be recalled that if an impedance $\underline{Z}_o$ is connected to the output terminals of a perfect transformer, the input impedance $\underline{Z}_i$ is $N^2\underline{Z}_o$ where N is the turns ratio. The "impedance level" is thus changed. At the beginning of the chapter it was stated that short sections of ultra-high-frequency transmission line may be used effectively for impedance matching. When so used, transmission-line losses are usually negligible, and theory based on the lossless line suffices. For the lossless line (R = 0, G = 0, $\alpha = 0$),

$$\underline{Z}_K = Z_K = \sqrt{\frac{L}{C}}\,\underline{/0^\circ} \text{ ohms} \tag{11-39}$$

$$= 276 \log_{10} \frac{D}{a} \text{ for open-wire pair} \tag{11-40}$$

$$= 138 \log_{10} \frac{b}{a} \text{ for concentric-tube line} \tag{11-41}$$

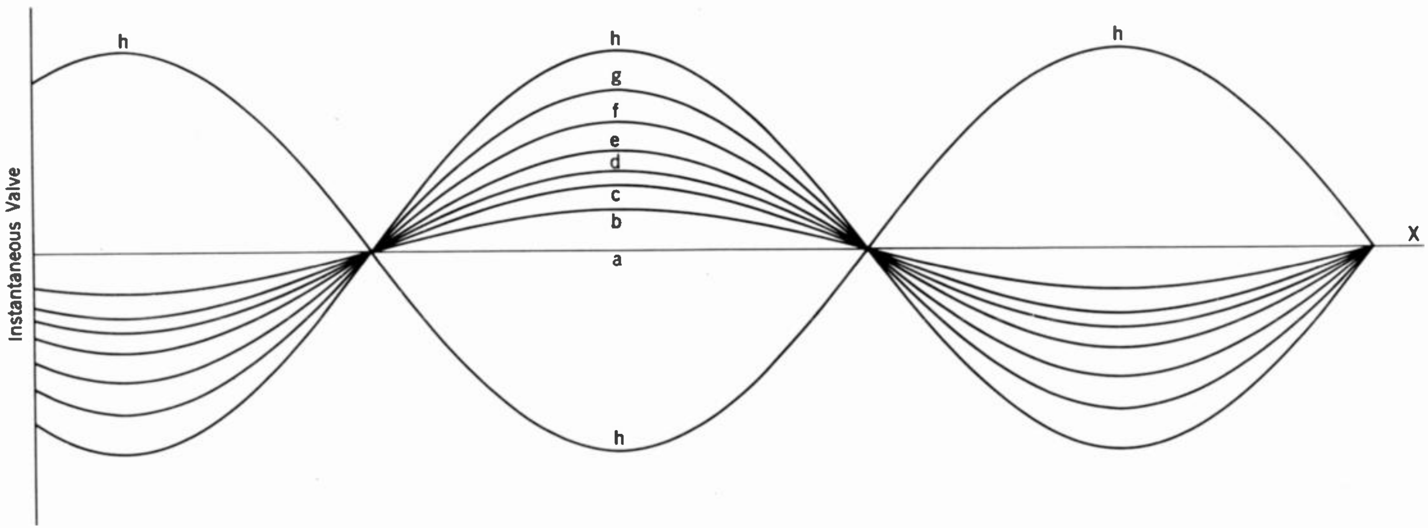


FIG. 11-6. Standing wave: a-h are for eight instants of time in first quarter-period. In next quarter-period wave changes from h back to a; in third quarter goes from a to h on opposite side of axis; in fourth quarter returns to a.

in which D is the axial spacing of the open-wire pair, a the radius of each wire, and b the inside radius of the concentric tube.

$$\underline{\gamma} = j\beta = j\omega\sqrt{LC} = j2\pi/\lambda \text{ numeric per unit length.} \qquad (11\text{-}42)$$

Table 11-I gives the general expression for the input or sending-end impedance $\underline{Z}_1$ for a line with output (load) impedance $\underline{Z}_0$ and by substituting in this the lossless condition, and also taking the value of $\underline{\gamma \ell}$ equal to $j\pi/2$ (one-quarter wavelength) the equation becomes

$$\underline{Z}_1 = \underline{Z}_K^2/\underline{Z}_0 \text{ complex ohms.} \qquad (11\text{-}43)$$

For the lossless-line case $\underline{Z}_K$ is real, and so if $\underline{Z}_0$ is a pure resistance, $\underline{Z}_1$ also will be resistive in character.

In order to match a source having an impedance $\underline{Z}_1$, to a load of impedance $\underline{Z}_0$, then, we may join them by means of a quarter-wave transmission line having a characteristic impedance $\underline{Z}_K$ equal to the geometric mean of the two impedances to be matched; in other words, $\underline{Z}_K = \sqrt{\underline{Z}_1 \underline{Z}_0}$.

The matching need not be exact, because most of the benefit to be had by properly matching a line is obtained by a reasonably close, though not exact, match.

<u>11-10. Impedance Matching by Use of a Stub Line.</u>- When a relatively long transmission line (e.g., several wavlengths) is feeding a load of impedance $\underline{Z}_0$, such as an untuned antenna having considerable reactance, matching of impedances may be accomplished by the use of a stub line branching out from the main line a short distance back from the load. The stub must be located at such a position on the main line, and have such an input impedance of its own, that the parallel impedance of the stub and the main line with termination $\underline{Z}_0$ shall be equal to $\underline{Z}_K$. Under this condition, reflections and standing waves are eliminated on the main line between the stub and the sending end, this being presumably much the longer part of the line.

Consider first the use of a short-circuited stub. This is usually preferable to an open-circuited one because of lower radiation losses and for purely mechanical reasons. We make use of the equation for $\underline{Z}_1$ from Table 11-I. It becomes for the lossless case, since $\underline{Z}_K = R_K$ and $\underline{\gamma} = j\beta$,

$$\underline{Z}_1 = R_K \frac{\underline{Z}_0 \cos \beta x_1 + j R_K \sin \beta x_1}{R_K \cos \beta x_1 + j \underline{Z}_0 \sin \beta x_1} \text{ complex ohms} \qquad (11\text{-}44)$$

which represents the complex impedance looking into the main line toward the receiving end, at a distance x_1 back from the receiving end.

At this point x_1 we connect a short-circuited stub line of length x_2, and having the same values of Z_K and β as the main line. From Table 11-I the input impedance of this stub is equal to ($\underline{Z}_o = 0$)

$$\underline{Z}_2 = jR_K \tan \beta x_2 \text{ ohms.} \qquad (11\text{-}45)$$

We may invert the expressions (11-44) and (11-45), writing them in the form of admittances Y_1 and Y_2. The admittance $\underline{Y}_2$ of the short-circuited stub of length x_2 is a pure susceptance, as might have been surmised in view of the assumption that there is no loss in the stub. The admittance of the parallel combination of stub and loaded line is $\underline{Y}_1 + \underline{Y}_2$, and the conductance, or real part, must clearly be that of $\underline{Y}_1$ alone, since $\underline{Y}_2$ contains no conductance term. For an impedance match, $\underline{Y}_1 + \underline{Y}_2$ must equal $1/R_K$, a real quantity, and so the value of x_1 must be selected first to make the real part of $\underline{Y}_1$ equal to $1/R_K$. Later the susceptances must be cancelled by selecting the proper length x_2 of the short-circuited stub. A graphic chart to assist in the calculation has been published by P.S. Carter.[8]

The two lengths may thus be worked out analytically, but as a practical matter the output (load) impedance $\underline{Z}_o$ is seldom known precisely, except by measurements made on the line feeding it. Suppose for example that the load is an antenna. Standing waves will be formed along the line feeding the antenna, unless the value of $\underline{Z}_o$ is equal to $\underline{Z}_K$, assumed here equal to $R_K + j0$. A voltage maximum point nearest the load may be located by means of a probe, and likewise a voltage minimum, which will be one-quarter wavelength away from the maximum. At both these points the impedance looking toward the load is a pure resistance; at the voltage maximum it is greater than R_K, and at the voltage minimum it is less than R_K. In the quarter wavelength toward the load from the voltage maximum, there is a position where the admittance looking toward the load has a real or conductance component equal to the reciprocal of R_K, and an imaginary or susceptance component which is negative or inductive. At this point we may connect a positive or capacitive susceptance of equal size so that the line from the generator "sees" a pure resistance equal to R_K at the junction. On the other side of the voltage maximum a similar point may be found at which the balance may be effected by the use of a shunt inductive susceptance. Measurements of current minimum and maximum will serve equally well, these occurring at voltage maximum and minimum respectively.

From Table 11-I it may be seen that the direct and reflected voltage waves have values respectively proportional to $(\underline{Z}_o + \underline{Z}_K)/2\underline{Z}_o$ and to $(\underline{Z}_o - \underline{Z}_K)/2\underline{Z}_o$. When they are in phase they

8. Charts for Transmission Line Measurements and Computations, RCA Review, 3, 355-68 (1938-39).

produce a voltage maximum, and when 180° out of phase a voltage minimum. The reflection coefficient of the load $\underline{Z}_o$ is

$$\frac{\underline{Z}_o - R_K}{\underline{Z}_o + R_K}, \tag{11-46}$$

this expression giving the phase shift ψ at reflection as well as the relative magnitude of the reflected wave. At any point x units distant from the load, a further shift in angle of $2\beta x = 4\pi x/\lambda$ will have taken place, since the phase of the incident wave is advanced, and that of reflected wave retarded as we go back from the load, each by an amount $\beta x = 2\pi x/\lambda$. Maxima will occur at points where $\psi + \beta x = \psi + 2\pi x/\lambda$ is equal to a multiple of 2π, and minima where this sum is an odd multiple of π.

The expressions for the stub locations and lengths are not simple, and the results are therefore presented in tabular form, Table 11-VI.

Table 11-VI

Location and Length of Stubs for Matching

$\frac{V_{max}}{V_{min}}$	Location (from V_{max} point)	Length of Stub	
		Stub Shorted	Stub Open
1.5	.14	.183	.067
2	.154	.150	.100
3	.170	.114	.136
4	.180	.092	.158
5	.188	.08	.170
6	.195	.071	.179
10	.205	.050	.200

All lengths are given in wavelengths, and the location is measured from a voltage maximum toward the generator if a short-circuited stub is to be used, from the voltage maximum toward the load if the stub is to be open. Some trial adjustment is usually necessary.

11-11. Line Parameters.- At ultra-high frequencies, the penetration of the current into the conductors is so small that the resistance is given very closely by the formula for effective depth of penetration. The effective depth of penetration δ is defined as the thickness of a conducting shell which would have the same resistance to direct current that the actual conductor has to alternating current. For standard annealed copper,

$$\delta = 2.61/\sqrt{f} \text{ inches} \tag{11-47}$$

and this holds very closely as long as the radius of curvature and the thickness of the conductor are everywhere large compared with δ. Based on this formula, the resistance of any conductor at ultra-high frequencies[9] is seen to vary as the square root of the frequency f, and for copper is equal to

$$\text{Resistance} = \frac{4.2\sqrt{\text{frequency}}}{\text{radius in cm}} \text{ microhms per meter,} \tag{11-48}$$

this expression applying to a single wire. For a concentric line, the expression also gives the resistance of the outer conductor, the inside radius of this conductor being then substituted in the denominator for "radius in cm."

The distributed shunt conductance G is not usually a very large source of loss, except when poor dielectric material is used for the beads which support the central conductor in a coaxial line. However, if the air is humid so that moisture condenses on the insulators the losses will be considerably increased. The beads give rise to small reflections, which also increase the losses to some extent. At exceedingly high frequencies, the loss in the insulation predominates because of the unavailability of really good dielectric materials for these frequencies.

The inductance of a concentric line at ultra-high frequencies is very closely

$$L = 0.4605 \log_{10} \frac{b}{a} \times 10^{-6} \text{ henry per meter,} \tag{11-49}$$

where b is the inside radius of the tube, and a the radius of the center conductor. The inductance due to linkages of the current with flux in the metal is negligible because of the small depth of penetration.

The capacitance of a concentric line is

$$C = \frac{0.241}{\log_{10} \frac{b}{a}} \times 10^{-10} \text{ farad per meter.} \tag{11-50}$$

The characteristic impedance is

$$\underline{Z}_0 = \sqrt{\frac{L}{C}} = 138 \log_{10} \frac{b}{a} \text{ ohms.} \tag{11-51}$$

For a balanced two-wire line, the corresponding parameters are, per meter of line,

$$R = \frac{8.4\sqrt{f}}{\text{radius in cm}} \text{ microhms per meter;} \tag{11-52}$$

9. A notable exception is found in a wave guide of circular cross section when operated in one mode or type of transmission, where the a-c resistance decreases indefinitely as the frequency is increased.

$$L = 0.921 \log_{10} \frac{D}{a} \times 10^{-6} \text{ henry per meter;} \qquad (11\text{-}53)$$

$$C = \frac{0.120}{\log_{10} \frac{D}{a}} \times 10^{-10} \text{ farad per meter;} \qquad (11\text{-}54)$$

$$\underline{Z}_0 = 276 \log_{10} \frac{D}{a} \text{ ohms.} \qquad (11\text{-}55)$$

Here D is the distance between centers of the two parallel wires each having radius a, measured in the same units as is D.

The resistance of a conductor at high frequencies is inversely proportional to its perimeter, other things being equal. The cost of a coaxial line, as well as the space it occupies, are determined principally by the tube size, and it is of interest to know what relative sizes of outer and inner conductors is the optimum under the condition of fixed outer conductor size. For lowest resistance, obviously the center conductor must be as large as possible; that is, its radius, a, should be just smaller than the inner radius of the tube, b. However, such a design would result in a very low value of inductance, high capacitance, and a low value of Q, defined as X/R. The conditions for the maximum value of Q may be computed readily, as follows:

$$Q = \frac{X}{R} = \frac{2\pi f L}{R}$$

$$= \frac{2\pi f \times 0.4605 \times 10^{-6} \log_{10} \frac{b}{a}}{4.2\sqrt{f}\,(\frac{1}{a} + \frac{1}{b})}$$

$$= \frac{0.70\, b(\text{cm}) \sqrt{f} \log_{10} \frac{b}{a}}{\frac{b}{a} + 1}, \qquad (11\text{-}56)$$

which has a maximum when b/a = 3.6. At this optimum ratio, and at 3000 megacycles per second, a coaxial line of one centimeter inside tube radius (b = 1 cm) would have a Q of 4650. This neglects any losses in the dielectric. Ratios of b/a lying between 2.5 and 7 may be used without increasing the attenuation more than 10 per cent above its minimum value for 3.6.

The characteristic impedance for a coaxial line having the optimum b/a of 3.6 is equal to 77 ohms.

11-12. Transients.- When a circuit produces a large gain through resonance effects, this gain is not manifest at once when the circuit is first energized. Instead, a period of growth of the output current and voltage amplitude takes place, with the duration of this period proportional to the Q of the elements forming the resonance circuit. Ordinarily, this transient period is of almost no practical importance in radio circuits, because the transient is usually substantially completed in a time which is insignificant in comparison with the duration of any of the signals. There are exceptions, however, and the simplest example of one might be in the handling of a very short pulse of high-frequency power repeated at intervals. If the amplification and generation of the high-frequency power depend largely on resonance effects, then the keying of the pulse at a low power level will not produce a square-wave pulse envelope, but instead one in which the leading "corner" of the square wave increases approximately exponentially with time toward a steady value.

Although limitations of space will not permit a thorough development of the transient theory, some consideration of the outstanding transient characteristics is important.

The basic differential equations of current and voltage on a transmission line are stated in (11-1) and (11-2) in terms of complex values, and these obviously become, for instantaneous values v and i of voltage and current respectively (see footnote 3 of this chapter),

$$\frac{\partial v}{\partial x} = Ri + L\frac{\partial i}{\partial t} \qquad (11\text{-}57)$$

and

$$\frac{\partial i}{\partial x} = Gv + C\frac{\partial v}{\partial t} \qquad (11\text{-}58)$$

Eliminating i to obtain an equation in v, x, and t; and then eliminating v to obtain an equation in i, x, and t, there result:

$$\frac{\partial^2 v}{\partial x^2} = LC\frac{\partial^2 v}{\partial t^2} + (RC + LG)\frac{\partial v}{\partial t} + RGv; \qquad (11\text{-}59)$$

$$\frac{\partial^2 i}{\partial x^2} = LC\frac{\partial^2 i}{\partial t^2} + (RC + LG)\frac{\partial i}{\partial t} + RGi. \qquad (11\text{-}60)$$

Actually during the transient period R varies considerably, and L to a smaller degree, owing to skin effect, and this variation had to be ignored to arrive at (11-59) and (11-60). Solutions for transients on lines are not in a convenient manageable quantitative form to include the phenomenon of skin effect, but the general behavior is not greatly affected thereby. Even when skin effect is neglected and all four line parameters are considered as constants, the solution is not simple, being expressed most conveniently for computation in an infinite series of Bessel functions.

The treatment here will be restricted to a brief description of the physical behavior of the transient. The impressed voltage variation will travel along the line with speed $1/\sqrt{LC}$, equal to the speed of light on an air-insulated line, and will undergo an attenuation proportional to the factor

$$e^{-\left(\frac{R}{2L} + \frac{G}{2C}\right)t} \tag{11-61}$$

and also some distortion which is a function of

$$\frac{R}{2L} - \frac{G}{2C} . \tag{11-62}$$

The distortion function is very complicated except when

$$R/L = G/C, \tag{11-63}$$

for which relation there is theoretically no distortion of the wave as it travels along the line (see Table 11-II). The steady state is arrived at by the accumulation of successive reflections; and if reflection is complete, as it is at the end of a short-circuited stub, the envelopes of the voltage or the current have a growth with time which is proportional to (11-61). The time constant of the growth function is the reciprocal of the quantity in parenthesis in (11-61). If $G = 0$, the time constant is $2L/R$. If the Q is large, the transient envelope would reach 63 per cent of its final value in Q/π cycles or periods of the alternating wave.

If the receiver end of the transmission line is not short-circuited or open-circuited, but has a resistance R_L, then only a partial reflection takes place when a wave front strikes the termination. In terms of voltage, the magnitude of the reflected wave bears to that of the incident wave a ratio equal to

$$\frac{\sqrt{C/L} - 1/R_L}{\sqrt{C/L} + 1/R_L} \tag{11-64}$$

This quantity is zero if $R_L = \sqrt{L/C}$; it is +1 if $R_L = \infty$ (open end) and -1 if $R_L = 0$ (short-circuited end). For other values of R_L the reflection coefficient (11-64) is a fraction, less than unity. When it is a fraction, the time constant of transient build-up is shortened to a new time proportional to the size of the fraction; that is, the time constant is the product of the fraction by the reciprocal of the parenthetical expression in (11-61).

BIBLIOGRAPHY

"The Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields," S.A. Schelkunoff, Bell System Tech. Jour., 13, 532-79 Oct. (1934).
"Communication Networks," vol. 2, E.A. Guillemin, New York, 1931.
"Electric Power Transmission and Distribution," 2nd ed., L.F. Woodruff, New York, 1938.

Chapter 12

RADIATION

In this chapter some of the simpler aspects of the theory of electromagnetic radiation are developed from the basic theory of electricity and magnetism; some attention is given to antennas for initiating, and for receiving, such radiation; and there is a limited discussion of the reflection, refraction, and diffraction of electromagnetic waves.

Commercial broadcasting is a commonplace matter with which virtually everyone is familiar. Frequencies used are of the order of magnitude of one megacycle. So-called short-wave broadcasting employs frequencies of the order of 10 megacycles; frequency-modulation broadcasts are assigned frequencies of the order of 45 megacycles; and television broadcasts, 100 megacycles. The following table gives an idea of the uses of various frequency channels in the United States as of 1940.

Table 12-I

Frequency Range in kc or Mc	Type of Service
100 - 200 kc	Mobile
160 - 370 kc	Aeronautical
370 - 515 kc	Mobile (also 370-380 kc maritime direction finding, 380-400 aeronautical)
515 - 550 kc	Government service
550 - 1600 kc	Broadcasting
1600 - 1750 kc	Mobile
1750 - 2050 kc	Amateur (other use in time of war)
2050 - 3485 kc	Mobile, police, aeronautical, etc., bands
3485 - 3500 kc	Experimental
3500 - 4000 kc	Amateur (other use in time of war)
4000 - 6000 kc	Mobile, aeronautical, etc., bands
6000 - 6200 kc	International broadcasting
6200 - 9500 kc	Mobile, aeronautical, amateur, etc., bands
9500 - 9700 kc	International broadcasting
9700 - 11700 kc	Fixed, and mobile, bands
11700 - 11900 kc	International broadcasting
11900 - 15100 kc	Fixed, mobile, amateur bands
15100 - 15350 kc	International broadcasting
25000 - 25275 kc	Facsimile

Table 12-I (Continued)

Frequency Range in kc or Mc	Type of Service
43 - 50 Mc	Frequency modulation broadcasting
50 - 108 Mc	Chiefly television, but includes government, amateur, aeronautical, airline navigation aid, etc., bands
108 - 300 Mc	Numerous
Above 300 Mc	Experimental, except 400-401 Mc amateur

Two questions immediately arise: (1) Why are such high frequencies used? and (2), If it is desirable to employ them, what limits their use?

A brief answer to the second question is that the most important limit on the use of high frequencies occurs because of the failure or insufficiency of generating and controling equipment at higher and higher frequencies. For example, in 1935 the use of 100 megacycles outside the laboratory would have been considered impractical; today frequencies of even higher magnitude are in practical use as a result of extensive research and development in the interim.

The first question -- Why are such high frequencies used? --involves several different aspects of communication: (1) effectiveness of radiation of electromagnetic energy increases with the frequency, radiation at very low frequencies being negligible for practical purposes; (2) each independent radio channel must use a frequency band (for example, 10 kilocycles in amplitude-modulation broadcasting; 2 megacycles or more in television broadcasting), and the simultaneous operation of many such channels means that the radiation from them must be at frequencies such that overlapping (interference) does not result; (3) the size of antennas and related equipment is smaller the greater the frequency -- this is particularly important in connection with portable equipment, including some military equipment; (4) in some respects, high frequencies are harder to "jam" than are lower frequencies--this matter may be of importance in military applications.[1]

12-1. A Change in Viewpoint.- In the preceding chapters the reader has been accustomed largely to deal with currents and voltages, and with other properties of circuits such as

1. Another reason for using high-frequency currents, not directly concerned with radiation however, is that certain circuits may transmit many messages simultaneously by properly adjusting frequencies; this is carrier telephony and may result in having as many as 200 channels on one transmission line.

resistance, inductance, and capacitance. 'In the more general aspects of electromagnetic theory, and particularly in connection with electromagnetic waves, it is convenient and often necessary to deal with electric and magnetic fields, and with properties of the medium in which the waves exist, such as conductivity, permeability, and dielectric constant or its equivalent. There is a rough parallelism which may be indicated thus:

Table 12-II

Item discussed in lumped-parameter circuit theory	Roughly analogous item discussed in electromagnetic wave theory
Voltages and currents at various branches of the circuit.	Electric and magnetic field intensities at points in the medium.
Resistance in a branch of the circuit.	Resistivity, or resistance of a unit cube, of the medium.
Inductance in a branch of the circuit.	Inductivity (permeability) of the medium = inductance of one turn of wire "per unit cube" of the medium when the medium is used as a core, and all flux concentrated in the core.
Capacitance in a branch of the circuit.	Capacitivity (dielectric constant) of the medium = capacitance of a capacitor in which unit cube of the medium is the dielectric between plates on two opposite faces.
Energy stored in a capacitor.	Energy stored in the electric field in a unit volume of the medium.
Energy stored in an inductor.	Energy stored in the magnetic field in a unit volume of the medium.

The above table is not intended to set up exact analogues, but rather to draw attention to the change in emphasis which will be evident in this chapter, and which is common to the study of electromagnetic radiation.

12-2. **System of Units to be Used.**- Table 12-III (pp. 372-374) has been constructed to recall various meanings and to

Table 12-III

QUANTITY	SYMBOL USED IN THIS CHAPTER	DESCRIPTION	NAME OF UNIT
Length, distance	s,r,x,y,z,	Fundamental (standard meter)	Meter
Area	S		Square meter = meter^2
Volume	τ		Cubic meter = meter^3
Time	t	Fundamental (astronomical basis)	Second
Velocity	c	Distance per unit time	Meter per sec
Acceleration	a	Velocity per unit time	(meter per sec)/sec = m/sec^2
Mass	-	Fundamental (standard kilogram)	Kilogram (despite the "kilo," this is the unit in the MKS system, not 1000 times the unit)
Force	-	Unit is force of attraction between two units of mass (kilograms) at points one unit of length (meter) apart times g	Unit of force = 10^5 dynes = 0.225 pound (sometimes called a newton)
Work, energy	-	Force times distance (work) or equivalent (energy)	Joule = 10^5 dyne-meters = newton-meter
Energy density	W	Energy per unit volume	Joule per meter^3
Power	P	Time rate of change of energy or time rate of doing work	Watt = joule per second
Power flux	$\wp$	Energy crossing unit area in unit time	Watt per meter^2
Power density	-	Time rate of change, or of dissipation, of energy per unit volume	Watt per meter^3

Current[1]	I	Fundamental (current balance)	Ampere
Current density	J	Current per unit area	Ampere per meter^2
Charge[1]	Q	Time integral of current	Coulomb = ampere-second
Charge volume density	ρ	Charge per unit volume	Coulomb per meter^3
Voltage, emf, difference of potential	V	Work per unit charge	Volt = joule per coulomb
Electric field intensity	E	Force per unit charge, or difference of potential per unit length	Volt per meter
Resistance	R	Voltage per unit current	ohm
Resistivity (specific[2])	-	Resistance of unit cube	ohm per meter cube = ohm meter
Conductance	G	Current per unit voltage	mho
Conductivity (specific[2])	-	Conductance of unit cube	mho per meter
Capacitance	C	Charge per unit voltage	Farad = coulomb per volt
Capacitivity (specific[2]) = dielectric constant	ε	Capacitance of capacitor using unit cube of material for dielectric	Farad per meter (see note 4)
Electric flux density	D	εE	Coulomb per meter^2
Electric flux	ψ	Electric flux density times area	Coulomb
Magnetomotive force[3]		Current through a surface (= work per unit pole)	Ampere-turn
Magnetic field intensity	H	Magnetomotive force per unit length (= force per unit pole)	Ampere-turn per meter
Magnetic flux	φ	Time integral of voltage (since time rate of change of φ is voltage per turn)	Weber = volt-second per turn = 10^8 emu (maxwells)

QUANTITY	SYMBOL USED IN THIS CHAPTER	DESCRIPTION	NAME OF UNIT
Magnetic flux density	B	Magnetic flux per unit area	Weber per meter^2
Inductance	L	Flux linkages per unit current	Henry = ohm-second
Inductivity (specific[2])	μ	Inductance, for one turn, per unit cube of material used as core	Henry per meter (see note 4)
Permeance	-	Flux per unit mmf (note close connection with inductance)	Henry per turn^2
Permeability (specific[2])	μ	Permeance per unit cube = B/H	Henry per meter (see note 4)

Notes: 1. It would have been more convenient to start the list of electric quantities with charge, but the practical fact that there is not a good physical standard of charge would have made such an arrangement unrealistic.

2. The word "specific" in the table indicates a quantity on a unit volume basis, and is sometimes used in the name, as specific conductance for conductivity. The word "specific" is also used in other senses; for example, to denote a dimensionless ratio such as specific inductive capacity, which is the ratio of the dielectric constant of any material to that of empty space. The specific quantities in the table all refer to properties of a material or medium.

3. By the law which states that the work required to carry a unit magnetic pole around any closed loop is equal to the current passing through any surface which has the loop for its periphery, it is seen that current (mmf) may be looked upon as work per unit pole corresponding to voltage (emf) as work per unit charge.

4. In the system of units outlined, μ for empty space is 1.26×10^{-6} henry per meter and ϵ is 8.85×10^{-12} farad per meter. The ratio $\sqrt{\mu/\epsilon}$ often appears; it has dimensions of impedance (compare $\sqrt{L/C}$) and is 377 ohms for empty space.

outline the system of units (MKS rationalized) which will be used. "Fundamental" quantities are those arbitrarily chosen as bases; all others are derived.

12-3. Displacement Current.- The current "through" the dielectric of a perfect capacitor is C dv/dt, and the current, other than leakage current, which "passes" from one wire of a two-wire transmission line to the other is likewise C dv/dt per unit length (C = capacitance per unit length). This current is not a conduction current corresponding to a flow of charge, but is a displacement current. Any varying electric field will produce a magnetic field, just as a conduction current will produce a magnetic field. The time rate of change of the electric flux ψ through any surface is called the displacement current through that surface because the magnetic field produced is directly proportional to it. In a capacitor $\psi = Cv$, hence the displacement current $d\psi/dt = C\ dv/dt$.

Displacement current and conduction current are indistinguishable as far as magnetic effects are concerned, but a displacement current is accompanied by no dissipation such as the RI^2 loss of conduction current in a conductor of finite conductivity.

It is essential that the reader have a good comprehension of displacement current, since it is of considerable importance in radiation. A simple law, such as the one which states that the work required to carry a unit magnetic pole around any closed path is equal to the total current flowing through any surface which has that path for its periphery, does not state that the work done is zero when the path is entirely in a perfect dielectric. Displacement current must be taken into account in determining the the total current in the broad interpretation of this relation appropriate to field and radiation problems.

12-4. Basic Laws.- As in any branch of a physical science, the basic laws of electromagnetic theory comprise those few laws from which the bulk of the theory--or at least a great part of it--can be derived. They represent the synthesis of the knowledge in the field. In dynamics the three Newtonian laws provide this basis, and in thermodynamics the two "laws of thermodynamics" are sufficient for most of the theory. In electromagnetic theory five basic laws constitute the nucleus of the theory, of which four, known as Maxwell's laws, will be required here. These four, which may at first glance seem relatively unimportant, and particularly so in comparison with the inverse square law, Ampere's law, etc., contain in their combination a tremendous amount of information, including the said inverse square law, Ampere's law, and most of the laws with which the

the reader is familiar in this field.

In their first rough form, the four basic laws needed here may be stated as follows:

I. The work required to carry a unit magnetic pole around any closed path once is equal to the total current (conduction and displacement) which is linked by that path.

II. The electric flux from a point charge q is equal in magnitude to q (some readers may be more accustomed to the statement that in electrostatic units the electric flux diverging from a point charge q is $4\pi q$).

III. The electromotive force induced in a closed loop fixed in position is the negative time rate of change of the magnetic flux linking that loop.

IV. Lines of magnetic flux are closed lines.

Using underscored symbols to denote vector quantities,[2] and the notation that a dot between two vectors indicates the triple product of the magnitude of the first vector, the magnitude of the second vector, and the cosine of the angle between them, the laws above may be expressed in the mathematical form

I. $\int_o \underline{H} \cdot d\underline{s} = I_{Total}$	o indicates any closed path, of which $d\underline{s}$ is a vector element; I_{Total} is the total of the conduction and displacement currents linking that path.
II. $\int_S \underline{D} \cdot d\underline{S} = Q$	S indicates any closed surface, Q is the total net charge in the volume bounded by the closed surface.
III. $-\frac{\partial \varphi}{\partial t} = e = \int_o \underline{E} \cdot d\underline{s}$	o indicates any stationary closed path, of which $d\underline{s}$ is a vector element; φ is the magnetic flux linking that closed path.
IV. $\int_S \underline{B} \cdot d\underline{S} = 0$	S indicates any closed surface of which a (vector) element is $d\underline{S}$.

12-5. Digression on Vectors.[3]- The mathematical form of the equations above is not well suited for present purposes. To

2. The terms "vector" and "vector quantity" as used in this chapter refer not to the vectors of a-c theory, which represent complex numbers, but to space vectors such as those representing simple mechanical forces, electric field intensity, etc.

3. The reader interested only in specific results may skip to the end of Sec. 12-14, where the intervening development is summarized. The material worked through up to that point is, however, useful not only in this chapter but also in Ch. 14--Hollow Guides.

obtain a better form, the equations will be applied to very small volumes and surfaces in the medium. This procedure is not necessary, and the equations could be used as they stand. But the increase in efficiency which will result from the present digression will amply repay the effort expended.

A magnetic field (physical concept) is a vector field (mathematical concept) in that at each point in the region where the field exists, and at each instant, the magnetic field intensity $\underline{H}$ has a definite magnitude and direction and the magnetic flux density $\underline{B}$ likewise has at each point at each instant a definite magnitude and direction. Thus $\underline{H}$ and $\underline{B}$, and similarly $\underline{E}$ and $\underline{D}$, are vector functions of position and time, that is, each vector depends on the coordinates xyz of the point at which it is to be measured or calculated, and on the time t. Consequently, the rate of change of a component of any one vector--$\underline{H}$ for example--with one variable must be indicated by a partial derivative. Thus

$$\frac{\partial H_y}{\partial x}$$

denotes the rate of change of H_y with x, while y, z, and t are held constant. Similarly $\partial H_y/\partial t$ denotes the rate of change of H_y with time t at a fixed point (xyz held constant).

Because $\underline{H}$, $\underline{B}$, $\underline{E}$, and $\underline{D}$ are functions of both position and time, it will be necessary to use partial derivatives rather extensively. It is highly desirable that the reader have a good physical picture of the process indicated by a partial derivative. Partial derivatives will come into this discussion shortly, meanwhile since we will be dealing with vectors rather extensively, we begin with some of the simpler aspects of vector manipulation. The following rules concerning space vectors (not related to the rotating vectors of a-c circuit theory, which represent complex numbers) are listed for review.

The geometrical definition of a vector is that a vector is a directed segment of a straight line. The direction is from the end called the origin toward the end called the terminus.

A positive scalar s times a vector $\underline{a}$ changes the length of the vector $\underline{a}$ from a to sa and leaves the direction unchanged.

Any vector $\underline{a}$ can be written as $\underline{a}_1 a$ where $\underline{a}_1$ is a vector of unit length parallel to $\underline{a}$ and a is the magnitude of $\underline{a}$.

The negative of $\underline{a}$ is $\underline{a}$ with its origin and terminus interchanged.

Two vectors are equal if they can be made to coincide by moving either one parallel to itself.

The sum $\underline{a} + \underline{b}$ of two vectors $\underline{a}$ and $\underline{b}$ is the vector from the origin of $\underline{a}$ to the terminus of $\underline{b}$ when the origin of $\underline{b}$ coincides with the terminus of $\underline{a}$. This definition can be extended to any number of vectors $\underline{a} + \underline{b} + \underline{c} + \ldots$, and furthermore, it can be shown by simple geometry that the order in which the vectors are added is immaterial.

To subtract $\underline{b}$ from $\underline{a}$, add $-\underline{b}$ to $\underline{a}$.

Any vector $\underline{a}$ can be written as the sum of a vector parallel to the X-axis, another parallel to the Y-axis, and a third parallel to the Z-axis where XYZ are the axes of a fixed, right-handed,[4] rectangular Cartesian coordinate system. If $\underline{i}$, $\underline{j}$, $\underline{k}$ are vectors of unit length parallel to X, Y, Z respectively then $\underline{a} = \underline{i}a_x + \underline{j}a_y + \underline{k}a_z$ where a_x, a_y, a_z are the magnitudes of the components of $\underline{a}$ parallel to X, Y, Z respectively.

The magnitude of $\underline{a}$ is $a = \sqrt{a_x^2 + a_y^2 + a_z^2}$.

If $\underline{a} = 0$, then a_x, a_y, a_z are zero.

In any vector equation each rectangular component on the left-hand side may be equated to the corresponding component on the right-hand side.

Thus far, some of the simpler vector rules have been reviewed. Extending now, the manipulation of vectors to multiplication, two types of multiplication are defined, The dot product is

$$\underline{a} \cdot \underline{b} = ab \cos (\underline{a}, \underline{b})$$

where cos ($\underline{a}$, $\underline{b}$) indicates the cosine of the angle between $\underline{a}$ and $\underline{b}$. Applying this definition to the unit vectors $\underline{i}$, $\underline{j}$, $\underline{k}$, it follows that $\underline{i} \cdot \underline{i} = \underline{j} \cdot \underline{j} = \underline{k} \cdot \underline{k} = 1$ since the angle between the vectors is 0° in each case, and $\underline{i} \cdot \underline{j} = \underline{j} \cdot \underline{k} = \underline{k} \cdot \underline{i} = 0$ since the angle between the vectors is 90° in each case. Hence if

$$\underline{a} = \underline{i}a_x + \underline{j}a_y + \underline{k}a_z \qquad (12\text{-}1)$$

and

$$\underline{b} = \underline{i}b_x + \underline{j}b_y + \underline{k}b_z \qquad (12\text{-}2)$$

then[5]

$$\underline{a} \cdot \underline{b} = a_x b_x + a_y b_y + a_z b_z = ab \cos (\underline{a}, \underline{b}) \qquad (12\text{-}3)$$

The cross product of two vectors, which may be chosen $\underline{E}$ and $\underline{H}$ because the cross product will be used later in connection with them, is defined

$$\underline{E} \times \underline{H} = \underline{\varepsilon}_1 EH \sin (\underline{E}, \underline{H}) \qquad (12\text{-}4)$$

and is the vector of magnitude EH sin ($\underline{E}$, $\underline{H}$) and direction perpendicular to $\underline{E}$ and $\underline{H}$ in the direction a right-hand screw would move if its head were placed in the plane of $\underline{E}$ and $\underline{H}$ and rotated from $\underline{E}$ toward $\underline{H}$ through the smaller angle between $\underline{E}$ and $\underline{H}$. The vector $\underline{\varepsilon}_1$ indicates a unit vector in this direction.

---

4. Three axes, X, Y, Z, form a right-handed system if, when the head of a right-handed screw is placed in the plane of any two and is rotated from the first in XYZXY toward the second through the smaller angle between the two, the screw moves in the positive direction of the third.
5. Assuming $\underline{a} \cdot (\underline{b} + \underline{c}) = \underline{a} \cdot \underline{b} + \underline{a} \cdot \underline{c}$, which the reader may show.

Applying this definition to the unit vectors $\underline{i}$, $\underline{j}$, $\underline{k}$, there result $\underline{i} \times \underline{j} = \underline{k}$, $\underline{j} \times \underline{k} = \underline{i}$, and $\underline{k} \times \underline{i} = \underline{j}$. It will at the same time be noticed that $\underline{i} \times \underline{j} = \underline{k}$, but $\underline{j} \times \underline{i} = -\underline{k}$, which illustrates a general rule that $\underline{E} \times \underline{H} = -\underline{H} \times \underline{E}$. That is, the definition of the cross product is such that one of the rules of common algebra, viz., that xy = yx, does not hold for it.

Now since

$$\underline{E} = \underline{i}E_x + \underline{j}E_y + \underline{k}E_z$$

and

$$\underline{H} = \underline{i}H_x + \underline{j}H_y + \underline{k}H_z$$

it follows[6] that

$$\underline{E} \times \underline{H} = \underline{i}\,(E_yH_z - E_zH_y) + \underline{j}(E_zH_x - E_xH_z) + \underline{k}(E_xH_y - E_yH_x) \qquad (12\text{-}5)$$

which gives the XYZ components of the vector $\underline{E} \times \underline{H}$.

Now consider the law (II) which states that the electric flux diverging from a charge Q is equal in magnitude to Q. Let S denote the whole of any closed surface, and let $d\underline{S}$ be a vector equal in magnitude to the area of any element of the surface and directed perpendicularly outward from the surface. Then $\underline{D} \cdot d\underline{S}$ is the electric flux crossing the element of surface $d\underline{S}$, and the total flux crossing S is

$$\int_S \underline{D} \cdot d\underline{S} \qquad (12\text{-}6)$$

where the integral is the sum of all the products $\underline{D} \cdot d\underline{S}$ for every element of surface of the closed surface S. The total charge Q inside S may be written

$$\int_\tau \rho d\tau \qquad (12\text{-}7)$$

where ρ is the charge density, $\rho d\tau$ the charge in any element of volume $d\tau$, and the whole volume τ is the volume bounded by the closed surface S. Thus the law (II) becomes

$$\int_S \underline{D} \cdot d\underline{S} = \int_\tau \rho d\tau \qquad (12\text{-}8)$$

We will now apply this to a very small (infinitesimal) volume $d\tau$ of the medium. The right-hand side will be simply $\rho d\tau$. The left-hand side will be the sum of $\underline{D} \cdot d\underline{S}$ for each of the six faces.

Following Fig. 12-1, for face 1234 $d\underline{S} = -\underline{i}\,dydz$, so that if D_x is the average value of the X component of $\underline{D}$ over 1234,

$$\underline{D} \cdot d\underline{S} = -D_x\,dydz \qquad \text{for 1234}$$

For 5678, $d\underline{S} = \underline{i}\,dydz$ and the average of the X component of $\underline{D}$ over it will be the average value of D_x over 1234 plus the rate of change of D_x in passing in the X direction $(\partial D_x/\partial x)$ times the distance (dx) between 1234 and 5678, whence

6. Assuming $\underline{a}$ in cross product with the vector $\underline{b} + \underline{c}$ is equal to $\underline{a} \times \underline{b} + \underline{a} \times \underline{c}$ which the reader may show.

$$\underline{D} \cdot d\underline{S} = \left(D_x + \frac{\partial D_x}{\partial x} dx\right) dydz \qquad \text{for } 5678$$

The sum of the $\underline{D} \cdot d\underline{S}$'s for 1234 and 5678 is

$$-D_x \, dydz + \left(D_x + \frac{\partial D_x}{\partial x} dx\right) dydz = \frac{\partial D_x}{\partial x} dxdydz$$

Likewise, the two faces perpendicular to Y (2187 and 6543) will contribute $(\partial D_y/\partial y)$ dxdydz and the two perpendicular to Z(1458 and 7632) will add $(\partial D_z/\partial z)$ dxdydz, so that in toto

$$\left(\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z}\right) dxdydz = \rho d\tau \tag{12-9}$$

But since dxdydz = $d\tau$, there finally results on a unit volume basis (dividing through by $d\tau$)

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \rho \qquad \text{coulombs} \tag{12-10}$$

which is our desired form. The right-hand side ρ is the charge in a unit volume, so the left-hand side must, by the law from which we started, be the net outward flux of $\underline{D}$. Thus (12-10) is a repetition of the law II, but in a form much more convenient for some uses.

If we define a vector operator del

$$\nabla \equiv \underline{i} \frac{\partial}{\partial x} + \underline{j} \frac{\partial}{\partial y} + \underline{k} \frac{\partial}{\partial z} \tag{12-11}$$

Then by the definition of a dot product

$$\nabla \cdot \underline{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \rho \tag{12-12}$$

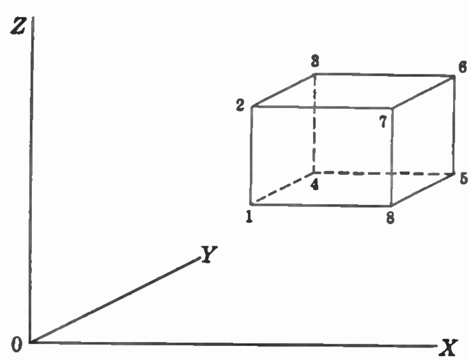


Fig. 12-1. Elementary parallelepiped laws I-IV of Sec. 12-4 are applied to obtain the differential forms of the laws.

or simply $\nabla \cdot \underline{D} = \rho$ coulombs (12-13)

and the operation $\nabla \cdot$ on any vector denotes the net outward flux of that vector per unit volume.

The law (IV) that $\underline{B}$ is continuous, that is, that as much magnetic flux enters as leaves any volume, may now be written

$$\nabla \cdot \underline{B} = 0 \qquad (12\text{-}14)$$

(The net outward flux of $\underline{B}$ from any unit volume = 0.)

Turn now to the law (I) of Sec. 12-4 which has been stated: the work required to carry a unit magnetic pole around any closed path is equal to the total current through any surface which has that closed path for its periphery; or in mathematical form

$$\int_o \underline{H} \cdot d\underline{s} = I_{Total} \qquad (12\text{-}15)$$

This law may be applied to the path 56785 around one side of the elementary parallelepiped of sides dx dy dy shown in Fig. 12-1.

Along 85, $d\underline{s} = \underline{j}dy$ and $\underline{H} \cdot d\underline{s} = H_y dy$ where H_y is the average of the values of the Y component of $\underline{H}$ along 85.

Along 67, $d\underline{s} = -\underline{j}dy$ and the average value of the Y component of $\underline{H}$ is the average along 85 plus the rate of change with z times the distance (dz) traveled over in going from 85 to 67. Hence

$$\underline{H} \cdot d\underline{s} \text{ along } 85 = H_y dy$$

$$\underline{H} \cdot d\underline{s} \text{ along } 67 = \left(H_y + \frac{\partial H_y}{\partial z} dz\right)(-dy)$$

$$\text{Total, 85 and 67} = -\frac{\partial H_y}{\partial z} dydz$$

Likewise, the total for sides 56 and 78 is $+(\partial H_z/\partial y)$ dydz. Consequently, the integral of $\underline{H} \cdot d\underline{s}$ about 56785 is

$$\int_{56785} \underline{H} \cdot d\underline{s} = \left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right) dydz \qquad (12\text{-}16)$$

According to the physical law (I), this is to equal the current over the surface 56785. If $\underline{J}_{Total}$ is the vector density of the total current (conduction plus displacement), then since 56785 is perpendicular to X, J_x dydz is the current through 56785. Therefore

$$\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right) dydz = J_x \, dy \, dz$$

or, on a unit area basis (divide by the area dydz of 56785)

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = J_x \qquad (12\text{-}17)$$

Likewise, applying the law (I) to a small area (43654) perpendicular to Y, and also to a small area (76327) perpendicular to Z,

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = J_y \tag{12-18}$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = J_z \tag{12-19}$$

The three equations can be written as a unit by taking the previously defined vector operator ∇ in cross product with $\underline{H}$ (replace in (12-5) components of $\underline{E}$ by the components of ∇):

$$\nabla \times \underline{H} = \underline{J}_{Total}$$

This is a vector equation, whose component equations are (12-17), (12-18), and (12-19).

In a similar manner, recalling that the emf in a nonmoving circuit is the integral of $\underline{E} \cdot \underline{ds}$, the third law--that minus the time rate of change of magnetic flux through a fixed circuit is equal to the emf induced in that circuit--may be written

$$\nabla \times \underline{E} = -\frac{\partial \underline{B}}{\partial t} \tag{12-20}$$

12-6. Forms of the Basic Equations.- Returning now to the physical problem, the forms of the basic equations previously given in Sec. 12-4, and those derived in the digression on vector analysis in Sec. 12-5, may be brought together as shown in Table 12-IV. Each entry on a given line says the same as any other entry on that line, but the forms are different, the one to be used depending on the application at hand.

12-7. Equations for Simple Radiation Problems.- The derivative form of the equations gives the basis for the study of electromagnetic radiation. To simplify the problem, the work which follows will, unless otherwise noted, be restricted to non-static electromagnetic fields which may exist in a medium which is

(a) homogeneous (a sample of the medium taken at any point is exactly like a sample taken at any other point)
(b) isotropic (same properties in all directions; not like wood with different properties along the grain and across the grain)
(c) a perfect insulator (no conduction current, i.e., $\underline{J} = 0$; this is the case of empty space and closely approximates the atmosphere provided electric charges--ions and electrons--are not present)
(d) free of electric charge ($\rho = 0$)
(e) such that $\underline{D} = \varepsilon\underline{E}$ and $\underline{B} = \mu\underline{H}$ where ε and μ are constants.-

Table 12-IV

Basic Laws of Electromagnetic Theory (Maxwell's Laws) in Three Different Forms

LAW	STATEMENT IN WORDS	INTEGRAL FORM	DERIVATIVE FORM
I	The work required to carry a unit magnetic pole around any closed path is equal to the total current (conduction and displacement) linking that path, i.e., the total current passing through any surface which has that path for its periphery.	$\int_o \underline{H} \cdot d\underline{s} = I_{Total}$ where o indicates any closed path, $d\underline{s}$ is a vector element of length of that path, and I_{Total} is the total current (conduction and displacement) linking that path. $\underline{H}$ is the vector magnetic field intensity, and the integral indicates that $\underline{H} \cdot d\underline{s}$ is to be calculated for each element of the path, and summed.	$\nabla \times \underline{H} = \underline{J}_{Total}$ where $\nabla = \underline{i}\frac{\partial}{\partial x} + \underline{j}\frac{\partial}{\partial y} + \underline{k}\frac{\partial}{\partial z}$ is a vector operator, $\nabla \times \underline{H}$ is the cross-product of ∇ and $\underline{H}$, and $\underline{J}_{Total}$ is the vector total current density (conduction and displacement).
II	The electric flux diverging from a charge Q is equal in magnitude to Q.	$\int_S \underline{D} \cdot d\underline{S} = Q$ where S is here any closed surface, $d\underline{S}$ is a vector element of S, $\underline{D}$ is the vector electric flux density, Q is the net charge within S, and the integral indicates that $\underline{D} \cdot d\underline{S}$ is to be calculated for each element of S, and summed.	$\nabla \cdot \underline{D} = \rho$ where $\nabla \cdot \underline{D}$ is the dot product of ∇ and $\underline{D}$, and ρ is the charge per unit volume (charge density). The equation states that at each and every point the electric flux diverging from a unit volume $(\nabla \cdot \underline{D})$ is equal to the charge in that volume (ρ).
III	The emf induced in any fixed closed loop is equal to minus the time rate of change of the magnetic flux φ through that loop. By emf is meant the work required to carry a unit charge around the loop.	$\int_o \underline{E} \cdot d\underline{s} = -\frac{\partial \varphi}{\partial t}$ where o, $d\underline{s}$, and the integral have meanings as in I, and the time rate of change of φ is written as a partial derivative to indicate that the loop does not move (xyz of each point of the loop remain fixed). $\underline{E}$ is the vector electric field intensity.	$\nabla \times \underline{E} = -\frac{\partial \underline{B}}{\partial t}$ where $\underline{B}$ is the vector magnetic flux density.
IV	Magnetic flux lines are continuous	$\int_S \underline{B} \cdot d\underline{S} = 0$ where S, $d\underline{S}$, and the integral have meanings as for II, and $\underline{B}$ is the vector magnetic flux density.	$\nabla \cdot \underline{B} = 0$ which states that at any and every point the net outward flux of $\underline{B}$ from a unit volume is zero (as much enters as leaves).

Under these conditions the Maxwell equations become

I. $\nabla \times \underline{H} = \varepsilon \frac{\partial \underline{E}}{\partial t}$
II. $\nabla \cdot \underline{E} = 0$
} since $\underline{J} = 0$, $\underline{D} = \varepsilon \underline{E}$ and ε is constant, and $\rho = 0$

III. $-\mu \frac{\partial \underline{H}}{\partial t} = \nabla \times \underline{E}$
IV. $\nabla \cdot \underline{H} = 0$
} since $\underline{B} = \mu \underline{H}$ and μ is constant

To determine the fields which may exist in the medium, it is desirable to solve for $\underline{E}$ and $\underline{H}$. Differentiate I with respect to t:

$$\nabla \times \frac{\partial \underline{H}}{\partial t} = \varepsilon \frac{\partial^2 \underline{E}}{\partial t^2}$$

(since $\frac{\partial}{\partial t} \nabla \times \underline{H} = \nabla \times \frac{\partial \underline{H}}{\partial t}$, the reader may show this by writing out the equality in component form)

Substitute for $\partial \underline{H}/\partial t$ from II:

$$-\nabla \times (\nabla \times \underline{E}) = \mu \varepsilon \frac{\partial^2 \underline{E}}{\partial t^2} \qquad (12\text{-}21)$$

Use in (12-21) the expansion[7]

$$\nabla \times (\nabla \times \underline{E}) = \nabla \nabla \cdot \underline{E} - \nabla^2 \underline{E} = -\nabla^2 \underline{E} \text{ by II;}$$

$$\nabla^2 \underline{E} = \frac{1}{c^2} \frac{\partial^2 \underline{E}}{\partial t^2} \qquad (12\text{-}22)$$

where $c = 1/\sqrt{\mu\varepsilon} = 3 \times 10^8$ meters per second ($\mu = 1.26 \times 10^{-6}$ henry per meter and $\varepsilon = 8.85 \times 10^{-12}$ farad per meter for empty space). The three components E_x, E_y, E_z must satisfy the three component equations of (12-22), namely

$$\frac{1}{c^2} \frac{\partial^2 E_x}{\partial t^2} = \nabla^2 E_x = \frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} \qquad (12\text{-}23)$$

$$\frac{1}{c^2} \frac{\partial^2 E_y}{\partial t^2} = \nabla^2 E_y \qquad (12\text{-}24)$$

$$\frac{1}{c^2} \frac{\partial^2 E_z}{\partial t^2} = \nabla^2 E_z \qquad (12\text{-}25)$$

Thus any electric field $\underline{E}$ which exists in the medium must satisfy (12-22), which is the same as saying that its components E_x, E_y, E_z must satisfy (12-23, 24, 25).

A similar derivation of $\underline{H}$ by eliminating $\underline{E}$ (differentiate III with respect to t, substitute for $\partial \underline{E}/\partial t$ from I) shows that

7. In general $\nabla \times \nabla \times \underline{V} = \nabla \nabla \cdot \underline{V} - \nabla^2 \underline{V}$, and this can be shown by expanding the two sides of the equation: write out $\nabla \times \underline{V}$ as in (12-17, 18, 19); write out $\nabla \times$ of the resultant vector, etc.

any field $\underline{H}$ which exists in the medium must satisfy

$$\nabla^2 \underline{H} = \frac{1}{c^2} \frac{\partial^2 \underline{H}}{\partial t^2} \tag{12-26}$$

12-8. **Special Case of Plane Waves.**- As the simplest case of interest, consider the field in which $\underline{E}$ depends on x and t only.[8] Then (12-23) reduces to

$$\frac{1}{c^2} \frac{\partial^2 E_x}{\partial t^2} = \frac{\partial^2 E_x}{\partial x^2} \tag{12-27}$$

(substitute E_y or E_z for E_x to obtain equations for other components)

This is a linear second order partial differential equation, similar in form to an equation met in transmission-line theory.[9] Its solution is

$$E_x = f_1(x - ct) + f_2(x + ct) \tag{12-28}$$

where f_1 and f_2 are any functions, and the notation $f_1(x - ct)$ indicates that (x - ct) must be used as a unit--the x and t parts cannot be separated. Thus $A(x - ct)^2$ might be f_1, but $A(x^2 - ct)$ could not be, since no use of (x - ct) taken as a unit will give $(x^2 - ct)$.

To show that (12-28) satisfies (12-27), let $\eta \equiv x - ct$; then $f_1(x - ct) = f_1(\eta)$ and

$$\frac{\partial f_1}{\partial x} = \frac{\partial f_1}{\partial \eta} \quad \text{and} \quad \frac{\partial^2 f_1}{\partial x^2} = \frac{\partial^2 f_1}{\partial \eta^2}$$

$$\frac{\partial f_1}{\partial t} = -c \frac{\partial f_1}{\partial \eta} \quad \text{and} \quad \frac{\partial^2 f_1}{\partial t^2} = c^2 \frac{\partial^2 f_1}{\partial \eta^2}$$

whence

$$\frac{\partial^2 f_1}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 f_1}{\partial t^2}$$

Likewise, $f_2(x + ct)$ satisfies (12-27), and since the equation is linear, the sum of the solutions f_1 and f_2 is a solution. It can be shown that (12-28) is the complete solution of (12-27).

Consider first the $f_1(x - ct)$ part of E_x; it states that that part of E_x will have the same value at all points on any plane perpendicular to the X-axis; it also states that if $x_1 - ct_1 = x_2 - ct_2$, then the value of E_x at the point x_1 at the time t_1 is equal to the value of E_x at x_2 at the time t_2. In other words, the value of E_x at x_1 at time t_1 has "traveled" to x_2 at the time t_2. Thus $f_1(x - ct)$ represents a wave, and since

---

8. Note that this does not mean that $\underline{E}$ has only an X component--all components of $\underline{E}$ may be present, each is a function of x and t only.
9. Actually, 12-27 is simpler than the equation which appears in footnotes 1 and 3 and in the last section of Ch. 11.

t_2 must be > t_1 (t_2 must be a later time than t_1) when $x_2 > x_1$ in order that $x_1 - ct_1 = x_2 - ct_2$, the wave travels in the + X direction. This wave is called the incident wave.

Likewise, it can be shown that $f_2(x + ct)$ represents a wave traveling in the -X direction, and called the reflected wave. The total solution (12-28) thus indicates that E_x may be made up of one resultant incident wave $f_1(x - ct)$, and one resultant reflected wave $f_2(x + ct)$. The solution does not require that both waves be present in any given case. The functions f_1 and f_2 in (12-28) are arbitrary; each may be present, or one or both may be zero.

Over any plane perpendicular to the direction of propagation (X), the vector electric field intensity due to the incident wave, or that due to the reflected wave, has the same vector value. For this reason the wave being considered in this special case is called a plane wave, and any plane perpendicular to the direction of propagation (X) is called a wave-front of the plane wave.

Plane waves are of importance because (1) they constitute the simplest type of wave from the standpoint of analysis, and (2) at large distances from the antenna or other radiating source, most waves approximate plane waves when considered over an area whose dimensions are small in comparison with the distance from the source. Figure 12-2 shows a plot of the distribution with x of a wave f(x - ct) for three instants. The reader should compare this with the figure of a wave on a transmission line, given in Ch. 11.

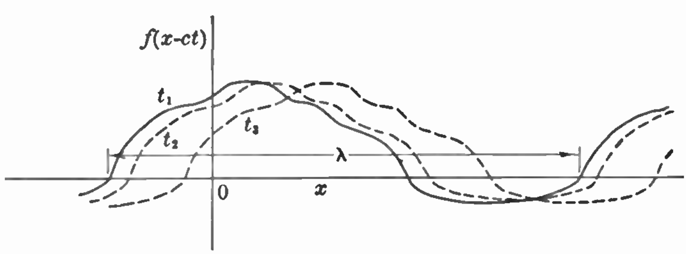


FIG. 12-2. Wave $f(x\text{-}ct)$ traveling in $+X$ direction, $(t_1 > t_2 > t_3)$ pictured at three instants t_1, t_2, t_3.

12-9. Transverseness of the Plane Wave.- The solution $E_x = f_1(x - ct) + f_2(x + ct)$ has counterparts in the solutions for E_y and E_z; these will be similar except that the arbitrary functions f_1 and f_2 will be replaced by other arbitrary functions (the functions appearing in E_y and E_z need not be the same as in E_x):

$$E_y = f_3(x - ct) + f_4(x + ct) \tag{12-29}$$

$$E_z = f_5(x - ct) + f_6(x + ct) \tag{12-30}$$

The E_x, E_y, E_z give the resultant $\underline{E}$, which must satisfy

$$0 = \nabla \cdot \underline{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \tag{12-31}$$

Substituting E_y and E_z from (12-29, 30), it follows that since they contain only x and t, derivatives with respect to y and z will be zero. Hence (12-31) reduces to

$$0 = \frac{\partial E_x}{\partial x}$$

which, for practical purposes here, states that $E_x = 0$. If an electromagnetic wave is such that $\underline{E}$ is a function of x and t only, then $\underline{E}$ has no component in the direction X in which the wave is propagated; the components E_y and E_z transverse to the direction of propagation may exist, hence $\underline{E}$ is a transverse electromagnetic wave.[10]

Likewise, as is more or less evident from the similarity of the controlling equations for $\underline{E}$ and $\underline{H}$, the magnetic field intensity $\underline{H}$ is also transverse in the case of a plane wave.

12-10. Relationships between E and H for Plane Wave.- When $\underline{E}$ is a function of x and t only (plane wave), E_x must be zero and equation III (Table 12-IV) gives

$$H_x = 0 \text{ and } \mu\frac{\partial H_y}{\partial t} = \frac{\partial E_z}{\partial x} \text{ and } \mu\frac{\partial H_z}{\partial t} = -\frac{\partial E_y}{\partial x} \tag{12-32}$$

which enable $\underline{H}$ to be calculated from $\underline{E}$. Conversely, if $\underline{H}$ is given, and is a function of x and t only (plane wave), then H_x must be zero and the components of $\underline{E}$ may be obtained from

$$E_x = 0 \text{ and } \varepsilon\frac{\partial E_y}{\partial t} = -\frac{\partial H_z}{\partial x} \text{ and } \varepsilon\frac{\partial E_z}{\partial t} = \frac{\partial H_y}{\partial x} \tag{12-33}$$

These illustrate the fact stated by I and III, namely that a non-static electric field is always accompanied by a magnetic field, and vice versa, and that when one is known the other may be determined.

If $\underline{E}$ is taken in the general form (12-29, 30) then $\underline{H}$ in terms of $\underline{E}$ is $H_x = 0$, $\mu c H_y = - f_5(x - ct) + f_6(x + ct)$, and

$$\mu c H_z = f_3(x - ct) - f_4(x + ct) \tag{12-34}$$

10. But not all electromagnetic waves are transverse; the proof here is limited to a plane wave.

These results may be checked by substituting in (12-32). Now $\underline{E} \cdot \underline{H} = E\,H \cos(\underline{E}, \underline{H}) = E_x H_x + E_y H_y + E_z H_z$ and substitution of (12-29, 30) and (12-34) show that for the incident wave alone, or for the reflected wave alone, $\underline{E} \cdot \underline{H} = 0$. That is, $\cos(\underline{E}, \underline{H})$ which is the cosine of the angle between $\underline{E}$ and $\underline{H}$, is zero. Consequently $\underline{E}$ and $\underline{H}$ are everywhere mutually perpendicular at such times that only the incident, or only the reflected, wave is present.[11]

12-11. Polarization of a Plane Wave.- It is usually convenient to consider the arbitrary functions in (12-29, 30) as cosines or the equivalent. Since the physical system--the medium through which the waves are being propagated--is linear,[12] the part of the resultant field varying at any one frequency is independent of the field at any other frequency. Consequently, a study of the field when a cosine function is assumed for f_1 will give a typical result. The argument here is similar to that used in lumped-parameter linear circuits, in which it is customary to study the effect of a sine-wave applied emf, on the assumption that one sine-wave applied emf gives a result typical of all, and that numerous such applied emf's give a result which is the sum of those due to each sine-wave.

Consider the particular case in which

$$E_y = A \cos \beta(x - ct) = A \cos (\omega t - \beta x)$$

and $$E_z = B \cos[\beta(x - ct) - \delta] = B \cos[\omega t - \beta x + \delta] \qquad (12\text{-}35)$$

where δ is a constant phase difference. Write w for $\omega t - \beta x$, and eliminate w between the two equations:

$$E_z = B(\cos w \cos \delta - \sin w \sin \delta)$$
$$= B\left(\frac{E_y \cos \delta}{A} - \sqrt{1 - \frac{E_y^2}{A^2}}\, \sin \delta\right)$$

or $$\frac{E_y^2}{A^2} + \frac{E_z^2}{B^2} - 2\,\frac{E_y E_z \cos \delta}{AB} = \sin^2 \delta \qquad (12\text{-}36)$$

This is the equation of an ellipse (plot E_z against E_y for various values of w). The wave (12-35) is said to be <u>elliptically polarized</u>.

If $\delta = \pi/2$ and $A = B$, eq. (12-36) reduces to

$$E_y^2 + E_z^2 = A^2$$

that is, to a circle. The wave (12-35) is then said to be <u>circularly polarized</u>.

---

11. It should be particularly noted that the proof applies <u>only to a plane wave</u>, and then only when the incident wave alone, or the reflected wave alone, is present.
12. A definition of a linear system is given in Ch. 1 and is applicable here.

If $\delta = 0$ or π, eq. (12-36) becomes the equation of a straight line and the wave is said to be linearly polarized.

The terms horizontal polarization and vertical polarization are often used; they indicate that the $\underline{E}$ vector is parallel to the horizon, or is vertical, respectively.

If instead of E_y and E_z, a voltage $e = A \cos \omega t$ and a current $i = B \cos(\omega t + \delta)$ were used, eq. (12-36) would become

$$\frac{e^2}{A^2} + \frac{i^2}{B^2} - \frac{2\, ei \cos \delta}{AB} = \sin^2 \delta \qquad (12\text{-}37)$$

which is the equation of the ellipse seen on a cathode-ray oscillograph when a voltage is applied to one pair of plates, and the drop caused by a current of the same frequency to the other.

12-12. Energy in the Fields, and the Propagation of Energy.- It will be recalled from elementary theory that in any region in which an electric field exists, there is an amount of energy $1/2\, \varepsilon E^2$ in each unit volume in the field ($1/2\, \varepsilon E^2$ per unit volume is similar to the $1/2\, Cv^2$ energy stored in the dielectric of a capacitor), and likewise in any magnetic field there is an amount of energy $1/2\, \mu H^2$ per unit volume (similar to the $1/2\, Li^2$ of an inductor), so that whenever the two fields $\underline{E}$ and $\underline{H}$ exist together, the total energy per unit volume is

$$\frac{1}{2}\varepsilon E^2 + \frac{1}{2}\mu H^2 \qquad \text{joules} \qquad (12\text{-}38)$$

The increase in this energy per unit time must equal, in an insulating medium (no dissipation, generation, or absorption), the rate at which energy is supplied to the unit volume. Conversely, the rate of decrease of the energy in a unit volume per unit time must equal the rate of divergence of energy from the unit volume; in mathematical form

$$-\frac{\partial}{\partial t}\left(\frac{1}{2}\varepsilon E^2 + \frac{1}{2}\mu H^2\right) = \nabla \cdot \underline{P} \qquad \text{watts} \qquad (12\text{-}39)$$

where $\underline{P}$ is a vector in the direction of energy flow at each point, and equal in magnitude to the energy crossing unit area normal to the direction of energy flow per unit time. It is the power flux of Table 12-III.

It is possible to find $\underline{P}$ in terms of $\underline{E}$ and $\underline{H}$ from the basic equations I and III: multiply the X component equation of I by E_x, the Y equation by E_y, and the Z by E_z, and add; this is the same as taking $\underline{E}$ in dot product with the vectors on the left, and on the right, side of I

$$\varepsilon \underline{E} \cdot \frac{\partial \underline{E}}{\partial t} = \underline{E} \cdot \nabla \times \underline{H} \qquad (12\text{-}40)$$

which is the same as

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\varepsilon E^2\right) = \underline{E} \cdot \nabla \times \underline{H} \qquad (12\text{-}41)$$

as can be seen by writing $E^2 = E_x^2 + E_y^2 + E_z^2$ and expanding to get (12-39).

Likewise, taking $\underline{H}$ in dot product with the vector on the left, and that on the right, side of III,

$$-\mu \underline{H} \cdot \frac{\partial H}{\partial t} = \underline{H} \cdot \nabla \times \underline{E}$$

or

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\mu H^2\right) = -\underline{H} \cdot \nabla \times \underline{E} \qquad (12\text{-}42)$$

Adding (12-41) and (12-42)

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\varepsilon E^2 + \frac{1}{2}\mu H^2\right) = \underline{H} \cdot \nabla \times \underline{E} - \underline{E} \cdot \nabla \times \underline{H} \qquad (12\text{-}43)$$

The two scalars on the right-hand side combine to give $-\nabla \cdot (\underline{E} \times \underline{H})$, which can be checked by forming $\underline{E} \times \underline{H}$ in component form (see eq. (12-5)), taking $\nabla \cdot$ of the resulting vector, and comparing this result with the scalar obtained by carrying out the indicated operation on the right-hand side of (12-43). Consequently,

$$-\frac{\partial}{\partial t}\left(\frac{1}{2}\varepsilon E^2 + \frac{1}{2}\mu H^2\right) = \nabla \cdot (\underline{E} \times \underline{H}) \qquad (12\text{-}44)$$

which says that the decrease in the electromagnetic energy in a unit volume per unit time is equal to $\nabla \cdot (\underline{E} \times \underline{H})$. Since it is also equal to $\nabla \cdot \underline{P}$, where $\underline{P}$ is the vector flow of energy across unit area in unit time.

$$\underline{P} = \underline{E} \times \underline{H} \qquad \text{watts per square meter} \qquad (12\text{-}45)$$

The vector $\underline{P}$ is known as Poynting's vector, and is extremely important in radiation and wave-guide theory. At each point in an electromagnetic field, it gives the direction of energy flow and is in magnitude equal to the energy crossing unit area in unit time.[13] If the energy alternately flows away from a source (such as a current in a wire) and returns without diminution, there is no net energy radiated; if energy flows away from a system and only part returns, the difference is permanently lost to the source and is said to have been <u>radiated</u>. (This assumes an insulating medium in which there is no dissipation.) When energy is radiated, there must be at least one region over which $\underline{P}$ is always directed away from the source.

13. The use of $\underline{P}$ is not restricted to radiation; $\underline{P}$ will, for example, give the magnitude and direction of energy flow at each point in the medium around a 60-cycle three-phase power line. Such a result is seldom needed in power work, since the <u>total</u> energy being transmitted can be computed from currents and voltages, and there is little interest in the distribution in space of the energy being transmitted.

For the type of medium specified in Sec. 12-7, the above discussion has been general. We now apply it specifically to the wave

$$\underline{E} = \underline{i}E_x + \underline{j}E_y + \underline{k}E_z = \underline{j}A\cos(\omega t - \beta x) \qquad (E_x = E_z = 0) \qquad (12\text{-}46)$$

for which, using III,

$$\underline{H} = \underline{i}H_x + \underline{j}H_y + \underline{k}H_z = \underline{k}\sqrt{\frac{\varepsilon}{\mu}}\, A\cos(\omega t - \beta x) \qquad (H_x = H_y = 0) \qquad (12\text{-}47)$$

This is a plane wave (the values of $\underline{E}$ and $\underline{H}$ are the same over any plane perpendicular to X, and any such plane is a wave-front), a transverse wave ($\underline{E}$ and $\underline{H}$ are perpendicular to X), and a linearly polarized wave ($\underline{E}$ is always parallel to Y, and lies in the XY plane, which is the plane of polarization), traveling in the +X direction.[14] The instantaneoous electric field energy per unit volume is

$$W_E = \frac{1}{2}\varepsilon E^2 = \frac{1}{2}\varepsilon A^2 \cos^2(\omega t - \beta x) \qquad (12\text{-}48)$$

and the instantaneous magnetic field energy per unit volume

$$W_H = \frac{1}{2}\mu H^2 = \frac{1}{2}\varepsilon A^2 \cos^2(\omega t - \beta x) \qquad (12\text{-}49)$$

whence, for this particular wave, $W_E = W_H$ at each point and at each instant. The energy flow is given by

$$\underline{P} = \underline{E} \times \underline{H} = (\underline{j} \times \underline{k})\, E_y H_z = \underline{i}\sqrt{\frac{\varepsilon}{\mu}}\, A^2\cos^2(\omega t - \beta x) \qquad (12\text{-}50)$$

$\underline{P}$ is in the X direction and is always positive, hence the given electromagnetic field is one in which energy is always being propagated in the +X direction (direction of $\underline{i}$). The instantaneous rate at which energy is crossing any unit area perpendicular to X is $\sqrt{(\varepsilon/\mu)}\, A^2 \cos^2(\omega t - \beta x)$; the average rate is $1/2\sqrt{(\varepsilon/\mu)}\, A^2$.

The fact that $\underline{P}$ is in the +X direction and always positive is important. It indicates that energy is always flowing in the +X direction. This energy is always traveling away from the source of the waves--it is thus being radiated away from the source. If $\underline{P}$ were alternately positive and negative, energy would be flowing away from the source part of the time, but back toward the source another part of the time. At low frequencies the energy stored in the magnetic field associated with an inductor ($1/2\, Li^2$), and in the electric field associated with a capacitor ($1/2\, Cv^2$), is of the latter type, and radiated energy is negligible for practical purposes. Thus Poynting's vector

14. The ratio H_z/E_y is $\sqrt{\mu/\varepsilon}$ and is sometimes called the impedance of the medium for plane waves; it is equal to 377 ohms. The corresponding admittance $\sqrt{\varepsilon/\mu}$ is 2.65 millimhos. Note from (12-46 and 12-47) that E_y and H_z are always in phase at a given point.

in the low-frequency case would show energy alternately flowing away from, and returning toward, the wires of the inductor or the plates of the capacitor.

12-13. Fields Set up by a Current in a Straight Wire.- Plane waves are convenient for illustration and important in their own right, but it is evident that a current in a straight wire is not likely to set up plane waves in the surrounding medium. There is a symmetry such that, at any point in a plane perpendicular to a straight current-carrying wire, we would expect to find $\underline{E}$ (also $\underline{H}$) equal in magnitude and in the same direction relative to the wire as at any other point in the plane, at the same distance from the wire.

A linear antenna[15] is a straight current-carrying wire, and the next problem is to determine the radiation from it, that is, the fields which will be set up in the surrounding medium due to it. The problem may be solved directly by specifying the current in the antenna, but this requires an extension of the theory not warranted here. Instead, we can obtain the desired result by a method of judiciously guessing the fields and showing that they fit the conditions.

Before proceeding, it may be well to inquire why the fields are not calculated to determine $\underline{H}$ by some familiar method such as applying Ampere's law in the form given in elementary texts. The reader has undoubtedly at one time or another used this method to find the magnetic field due to a current in a single straight wire, and has obtained a result which (1) holds only at low frequencies, and (2) probably did not indicate the concomitant electric field. For results applicable at high frequenices, it is necessary to find the fields more accurately, as we will now proceed to do.

The problem may be formulated thus: we seek fields $\underline{E}$ and $\underline{H}$ which satisfy the following conditions:

1. They must satisfy the Maxwell equations, or their equivalents for present purposes

$$\nabla^2 \underline{E} = \frac{1}{c^2}\frac{\partial^2 \underline{E}}{\partial t^2} \quad \text{and} \quad \nabla \cdot \underline{E} = 0 \qquad (12\text{-}51)$$

and

$$\nabla^2 \underline{H} = \frac{1}{c^2}\frac{\partial^2 \underline{H}}{\partial t^2} \quad \text{and} \quad \nabla \cdot \underline{H} = 0; \qquad (12\text{-}52)$$

2. They must be such that $\underline{E}$ (likewise $\underline{H}$) has the same magnitude, and the same direction relative to the antenna, at

15. Rigorously, "An antenna is a conductor or a system of conductors for radiating or receiving radio waves, exclusive of the means for connecting its main portion with the associated apparatus," I.R.E. Standards, definition 2T1. A linear antenna is a straight wire of negligible diameter.

all points equidistant from the antenna in a plane perpendicular to it;

3. The magnetic lines of force (lines everywhere parallel to $\underline{H}$) must be circles lying in planes perpendicular to the antenna (this will be recognized as a low-frequency rule and holds equally well here);

4. The fields should be such that at low frequencies they reduce to known fields; in this way we can relate the fields determined under conditions (1), (2), and (3) with the source of the fields and, in particular, can try to show that certain fields are those due to the current in the linear antenna.

To make the problem definite, take the Z-axis along the antenna, choose the origin O on the antenna, let $r^2 \equiv x^2 + y^2 + z^2$ be the square of the distance from O to any point Q whose rectangular coordinates are xyz, and let θ be the angle from +Z to OQ ($90^\circ - \theta$ is called the elevation angle, or simply the elevation, when Z is vertical). See Fig. 12-3.

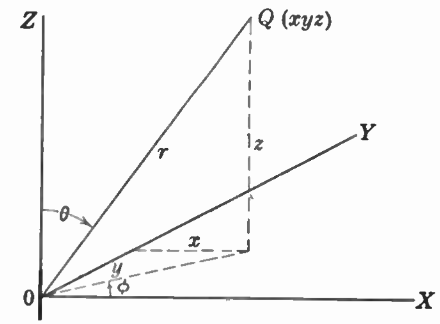


FIG. 12-3. Antenna (heavy line at origin O), any point Q whose coordinates are xyz in the XYZ rectangular Cartesian coordinate system, or, $r\theta\phi$ in the spherical coordinate system indicated.

By condition (3), $\underline{H}$ must be perpendicular to Z, that is, to the unit vector $\underline{k}$, and by the fourth Maxwell equation $\nabla \cdot \underline{H}$ must be zero. If φ is any scalar function of position (xyz) and time (t), $\underline{k}\varphi$ is a vector, and $\nabla \times (\underline{k}\varphi)$ is perpendicular to $\underline{k}$ (when expanded $\nabla \times (\underline{k}\varphi)$ has no component in the $\underline{k}$ direction). Furthermore, a simple expansion shows that $\nabla \cdot \nabla \times (\underline{k}\varphi)$, that is, $\nabla \cdot$ of the vector $\nabla \times (\underline{k}\varphi)$, is zero; thus, if we take for a trial

$$\underline{H} = \nabla \times (\underline{k}\,\varphi) \qquad (12\text{-}53)$$

the requirement $\nabla \cdot \underline{H} = 0$ will be fully satisfied and condition (3) will be partly satisfied. It will be fully satisfied, if, instead of being any function of xyz and t, φ is a function of r and t only. That is, we want φ to be expressible in terms of r and t alone, without having any x, y, or z appearing explicitly.

To show that if

$$\varphi = \varphi(r, t \text{ only}) \qquad (12\text{-}54)$$

then $\underline{H}$ satisfies condition (3), note that

$$\underline{H} = \nabla \times (\underline{k}\varphi) = \underline{i}\frac{\partial \varphi}{\partial y} - \underline{j}\frac{\partial \varphi}{\partial x} \qquad (12\text{-}55)$$

and when (12-54) is satisfied

$$\underline{H} = \underline{i}\,\frac{d\varphi}{dr}\frac{\partial r}{\partial y} - \underline{j}\,\frac{d\varphi}{dr}\frac{\partial r}{\partial x} = (\underline{i}y - \underline{j}x)\,\frac{1}{r}\frac{d\varphi}{dr}$$

$$= (\underline{r} \times \underline{k})\,\frac{1}{r}\frac{d\varphi}{dr} = (\underline{r}_1 \times \underline{k})\,\frac{d\varphi}{dr} = -(\underline{k} \times \underline{r}_1)\,\frac{d\varphi}{dr} \qquad (12\text{-}56)$$

where $\underline{r} \equiv \underline{i}x + \underline{j}y + \underline{k}z$ is the vector from 0 to any point Q whose coordinates are xyz, and $\underline{r}_1$ is a unit vector in the direction of $\underline{r}$. Now $\underline{r}_1 \times \underline{k}$ has a magnitude $\sin\theta$, since θ is the angle between $\underline{k}$ and $\underline{r}_1$, and has a direction perpendicular to $\underline{k}$ and $\underline{r}$. It is thus tangent to a circle, in a plane perpendicular to Z, passing through Q and with center on the Z-axis. As Q moves in this plane at a constant distance from Z, $\underline{H}$ remains tangent to the same circle and its magnitude remains fixed (both r and $d\varphi/dr$ remain fixed since r is unchanged). Hence, when φ is restricted according to (12-54), $\underline{H}$ is given by (12-56) and satisfies condition (3), and incidentally, also satisfies condition (2).

Another requirement on $\underline{H}$ is that $c^2\nabla^2\underline{H} = \partial^2\underline{H}/\partial t^2$. For $\underline{H}$ given by (12-56), it will be sufficient that

$$\nabla^2\varphi = \frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} \qquad (12\text{-}57)$$

because then $\partial\varphi/\partial y$ and $-\partial\varphi/\partial x$, that is, H_x and H_y, will satisfy (12-52).

Since φ now depends on r and t, a solution as simple as (12-28) cannot be obtained. However, if f_1 and f_2 are any functions,

$$\varphi = \frac{1}{r}\,f_1(r - ct) + \frac{1}{r}\,f_2(r + ct) \qquad (12\text{-}58)$$

is the solution, as can be shown by direct substitution. We will arbitrarily take the case in which $f_1 = A\cos\beta(r - ct)$ and $f_2 = 0$, so that

$$\varphi = \frac{A}{r}\cos\beta\,(r - ct) = \frac{A}{r}\cos(\omega t - \beta r) \qquad (12\text{-}59)$$

and it will be sufficient if the reader is convinced by substituting in (12-57) that this is a solution of (12-57).

For φ as just given, (12-59), equation (12-56) gives for the magnitude H of $\underline{H}$

$$H = A\sin\theta\left[\frac{\cos(\omega t - \beta r)}{r^2} - \frac{\beta\sin(\omega t - \beta r)}{r}\right] \qquad (12\text{-}60)$$

$$= \frac{A\sin\theta}{r^2}\sqrt{1 + \beta^2 r^2}\,\cos(\omega t - \beta r + \delta) \qquad (12\text{-}61)$$

where $\delta = \tan^{-1}(\beta r)$, and H is in ampere-turns per meter.

To summarize, this section was begun by formulating a problem, next an expression for the $\underline{H}$ field was obtained by guessing at a solution that would satisfy one of the conditions originally set down, and finally a complete expression was had by successively modifying the assumed solution until all conditions were satisfied. We now have a field $\underline{H}$, which is a physically possible field (satisfies Maxwell's equations) and which also must be closely related to the field (local plus radiation) of a current in a wire.

To investigate further, assume the low-frequency case: frequency $f = \omega/2\pi$ small, period T large, wavelength λ large, $\beta = 2\pi/\lambda$ small, and look at the field in a region in which $r \ll \lambda$ (at 60 cps, $\lambda = 3000$ miles in empty space; the region $r \ll \lambda$ includes virtually all the space in which a field measurement might be made). Since β is very small and $r \ll \lambda$, βr will be negligible for practical purposes, $\cos(\omega t - \beta r)$ becomes $\cos \omega t$, and the second term of (12-60) is negligible, hence

$$H = (A \cos \omega t) \frac{\sin \theta}{r^2} \qquad (12\text{-}62)$$

or

$$\underline{H} = \frac{\underline{k} \times \underline{r}_1}{r^2} (A \cos \omega t)$$

This is a low-frequency case, and in such a case Ampere's law in its usual form states that the field $\underline{H}$ due to a current i in a length of wire ds lying in the Z ($\underline{k}$) direction is

$$\underline{H} = \frac{\underline{k} \times \underline{r}_1}{4\pi r^2}\, i\, ds \qquad (12\text{-}63)$$

The field $\underline{H}$ given by (12-63) and that given by (12-62) are the same provided

$$i\, ds/4\pi = A \cos \omega t \qquad (12\text{-}64)$$

In other words, we surmise that the field (12-60) is the whole magnetic field at any frequency due to a current $i = I \cos \omega t$ in a length ds of wire lying in the Z direction, provided A is replaced by I ds/4π. The first term on the right-hand side is the field usually considered at low frequencies--it is called the local or the induction field. At high frequencies this field is important near the antenna ($r < 2\lambda$), but at distances far from the antenna ($r \gg \lambda$), the first term is negligible,[16] and

$$\underline{H} = (\underline{k} \times \underline{r}_1) \frac{A \beta \sin (\omega t - \beta r)}{r} \qquad (12\text{-}65)$$

16. It is well to emphasize that near the antenna ($r < 2\lambda$), the local field is important, and measurements in this region will not give the radiation field, irrespective of the frequency.

$$H = \frac{A\omega \sin\theta}{cr} \sin(\omega t - \beta r) = 0.021\ Af\,\frac{\sin\theta}{r} \sin(\omega t - \beta r) \quad (12\text{-}66)$$

where f is in Mc, the field thus defined is termed the radiation field; it decreases with the distance as $1/r$; it varies with elevation ($90^{\circ} - \theta$) from a maximum in the $\theta = \pi/2$ direction to zero in the direction coinciding with the antenna ($\theta = 0$).

The electric field $\underline{E}$ which is the necessary accompaniment of the $\underline{H}$ field given by (12-65) can be obtained by using the first Maxwell equation $\varepsilon\, \partial \underline{E}/\partial t = \nabla \times \underline{H}$ in component form. Thus

$$\varepsilon \frac{\partial E_x}{\partial t} = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} \quad (12\text{-}67)$$

and since H_x, H_y, and H_z are specified by (12-55), these can be inserted on the right-hand side, and the resulting equation for $\partial E_x/\partial t$ can be solved for E_x. Likewise E_y and E_z can be found. The results are slightly complicated and are given in a footnote.[17] We, however, can combine E_x, E_y, and E_z by projecting each component in the $\underline{r}_1$ direction, and in a direction $\underline{\theta}_1$ where $\underline{\theta}_1$ is a unit vector (unit magnitude or length, not unit angle) in a direction perpendicular to $\underline{r}$ in the $\underline{k}$, $\underline{r}$ plane, as shown in Fig. 12-4. This direction ($\underline{\theta}_1$) is called the "θ" direction because it is the direction in which the point Q would move if θ were increased slightly. The component E_r of $\underline{E}$ in the $\underline{r}_1$ direction is, recalling that the dot product of two vectors is the scalar equal to the product of the magnitude of one by the magnitude of the component of the other parallel to the first,

$$E_r = \underline{r}_1 \cdot \underline{E} = \frac{\underline{r} \cdot \underline{E}}{r} = \frac{1}{r}(xE_x + yE_y + zE_z) \quad (12\text{-}68)$$

and the component E_θ of $\underline{E}$ in the $\underline{\theta}_1$ direction can be obtained from

---

17.

$$E_x = x\,z\,U \qquad H_x = y\,V$$
$$E_y = y\,z\,U \qquad H_y = -x\,V$$
$$E_z = z^2 U - r^2(U + U') \qquad H_z = 0$$

where $U' = \dfrac{-2}{\beta r^5}\sqrt{\dfrac{\mu}{\varepsilon}}\,(\sin w + \beta r \cos w)$

$$U = \frac{1}{\beta r^5}\sqrt{\frac{\mu}{\varepsilon}}\,(\beta^2 r^2 - 3) \sin w + 3\beta r \cos w$$

$$V = (\cos w - \beta r \sin w)/r^3 = \frac{1}{r^3}\sqrt{1 + \beta^2 r^2}\,\sin(w + \delta') \text{ where } \delta' \equiv \tan^{-1}(\beta r)$$

$$w = \omega t - \beta r \text{ and } \sqrt{\mu/\varepsilon} = 377 \text{ ohms.}$$

It follows that $E_x H_x + E_y H_y + E_z H_z = 0$, that is, that $\underline{E}$ is perpendicular to $\underline{H}$. Thus $\underline{E}$ at every point lies in a plane containing the linear radiator, whereas $\underline{H}$ at every point is perpendicular to this plane.

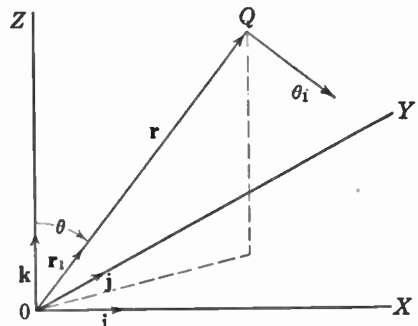


FIG. 12-4. Diagram showing the unit vector **i j k** and also $\mathbf{r}_1$ and θ_1. Note that $\mathbf{r}_1$ and θ_1 are in fixed directions, but $\mathbf{r}_1$ and θ_1 change directions in going from point Q to any other point.

$$E_{\epsilon}^2 + E_r^2 = E^2 = E_x^2 + E_y^2 + E_z^2 \qquad (12\text{-}69)$$

Relegating details to a footnote,[18] the vector electric field $\underline{E}$ at any point whose distance from 0 is r and whose elevation is $90^\circ - \theta$, has components

$$E_r = \frac{2\ A \cos\theta}{\varepsilon\,\omega\,r^3}\,[\sin(\omega t - \beta r) + \beta r \cos(\omega t - \beta r)] \qquad (12\text{-}70)$$

$$E_\theta = \frac{A \sin\theta}{\varepsilon\omega r^3}\,[\sin(\omega t - \beta r) + \beta r \cos(\omega t - \beta r) - \beta^2 r^2 \sin(\omega t - \beta r)] \qquad (12\text{-}71)$$

At low frequencies, e.g., 60 cps, the terms with $\beta(= 2\pi/\lambda = \omega/c)$ in them will be negligible for all values of r of practical interest (λ = 3000 miles for f = 60 cps, hence r will be much less than λ). At high frequencies (f large, λ small) the only term which will be important when r is appreciably > λ (for f = 300 megacycles, λ = 1 meter, hence the practical working region is in the range r > λ) is

$$E_\theta = \frac{A\,\mu\,\omega\,\sin\theta}{r}\,\sin(\omega t - \beta r) = \sqrt{\frac{\mu}{\varepsilon}}\,H = 377\ H \qquad (12\text{-}72)$$

This is the radiation electric field corresponding to the radiation magnetic field H.

18. From the previous footnote, $xE_x + yE_y + zE_z = (x^2 + y^2 + z^2)\,zU - zr^2U' = -\,zr^2U'$, whence $E_r = -\,zrU' = -\,r^2U' \cos\theta$. Likewise, $E_x^2 + E_y^2 + E_z^2 - E_r^2 = r^2z^2U^2 + r^4(U + U')^2 - 2\,r^2z^2U(U + U') + r^2z^2U'^2 = r^2(r^2 - z^2)(U + U')^2$; whence $E_\theta = r^2 \sin\theta\,(U + U')$.

We may stop to observe one fact which may help to bring out the distinction between the theory here developed and low-frequency theory. If we form Poynting's vector $\underline{P} = \underline{E} \times \underline{H}$, we get the energy crossing unit area in unit time at any point. Using the general forms of $\underline{E}$ (equations 12-70 and 71) and $\underline{H}$ (equation 12-60), the development[19] will show that every term alternates in sign except one. Thus the energy surges away from and back toward the source, and is not radiated, except for the energy corresponding to the one term. At low frequencies the terms alternating in sign are important in virtually all regions of practical interest, and the term of fixed sign is negligible. Hence, at low frequencies (e.g., 60 cps) there is no practical error in neglecting radiation. On the other hand at high frequencies the dominant term in $\underline{P}$ is, for $r \gg \lambda$,

$$\underline{P} = \underline{r}_1 \sqrt{\frac{\mu}{\varepsilon}}\; H^2 = \underline{r}\; \frac{cA^2 \omega^2 \sin^2 \theta}{\varepsilon\, r^2} \sin^2(\omega t - \beta r) \qquad (12\text{-}73)$$

which is simply the vector $\underline{E} \times \underline{H}$ using only the radiation fields. At high frequencies, near the source both the local and the radiation fields are present, energy corresponding to the former surging out from, and back toward, the source, and energy corresponding to the latter moving away from the source; at high

---

19. The vector $\underline{P}$, giving the instantaneous direction of energy flow at any point and the instantaneous rate at which energy at the point is crossing unit area perpendicular to the direction of flow, must be perpendicular to $\underline{H}$ everywhere (since $\underline{P} = \underline{E} \times \underline{H}$). Since we have already shown that $\underline{E}$ is perpendicular to $\underline{H}$, $\underline{P}$ and $\underline{E}$ must lie in the same plane, and are at right angles to one another in this plane, which is the plane containing the linear antenna. Thus $\underline{P}$, like $\underline{E}$, may be resolved into two components, one P_r parallel to $\underline{r}_1$ and one P_θ parallel to $\underline{\theta}_1$. By recalling that $\underline{r}_1$, $\underline{\theta}_1$, and $\underline{\varphi}_1$ are mutually perpendicular at each point, although they may vary in direction from point to point, it may be recognized that $\underline{\varphi}_1 \times \underline{r}_1 = \underline{\theta}_1$ and $\underline{\varphi}_1 \times \underline{\theta}_1 = -\underline{r}_1$. Thus, $\underline{P} = \underline{E} \times \underline{H} = (\underline{r}_1 E_r + \underline{\theta}_1 E_\theta) \times \underline{\varphi}_1 H = -\underline{\theta}_1 E_r H + \underline{r}_1 E_\theta H$; whence,

$$P_r = A' \left(P_1 + \beta^3 r^3 \sin^2 w\right) \sin^2\theta$$

$$P_\theta = -A' P_2 \sin\theta \cos\theta$$

where

$$P_1 = 1 - 2\beta^2 r^2 \sin 2w + \beta r \cos 2w$$

$$P_2 = 1 - \beta^2 r^2 \sin 2w + \beta r \cos 2w$$

$$A' = \frac{A^2}{2\beta r^5}\frac{\mu}{\varepsilon} = \frac{189\, A^2}{\beta r^5} \doteq 3\,\frac{\lambda A^2}{r^5}$$

$$w = \omega t - \beta r$$

If we stop at any point Q, thus fixing r and θ, and observe the variation of $\underline{P}$ with time t, it is seen that every term in P_1 and P_2 is periodic--alternately positive and negative--and only the term $A'\beta^3 r^3 \sin^2 w \sin^2\theta = 0.082 \frac{A^2 f^2}{r^2} \sin^2\theta \sin^2 w$ has an integral over a period which is different from zero.

frequencies and at distances $r \gg \lambda$ (most of the practical working region is included), the radiated energy alone is present for practical purposes.

12-14. **Summary of the Preceding Development.**- Starting with four fundamental laws of electromagnetic theory, and the further fact that a varying electric field (displacement current) produces a magnetic field just as a conduction current produces a magnetic field, it has been possible to outline the magnetic and electric fields in the medium about a very short, thin straight wire carrying a current. These fields may be considered as the sum of local fields, which are the only fields of importance at low frequencies; and radiation fields, which are important here. Low-frequency laws such as the usual elementary form of Ampere's law give only the local fields and consequently omit consideration of radiation. Here the process has been different: general results for the fields have been derived from which either the low-frequency or the high-frequency (radiation) parts can be obtained.

Assuming the antenna to be on the Z-axis at O in Fig. 12-3, the vector electric field intensity $\underline{E}$ at any point Q distant r from the antenna and at elevation $90^\circ - \theta$ is, in volts per meter,

$$\underline{E} = \underline{\theta}_1 \left(7.9 \, \frac{Af \sin\theta}{r}\right) \sin(\omega t - \beta r) \quad (f \text{ in Mc}) \quad (12\text{-}74)$$

provided $r \gg \lambda$. This is the radiation electric field, and the direction of $\underline{E}$ is in the direction of the unit vector $\underline{\theta}_1$, as shown on Fig. 12-4. Likewise, the vector magnetic field intensity $\underline{H}$ at any point Q is given under the same conditions by

$$\underline{H} = \underline{\phi}_1 \left(0.021 \, \frac{Af \sin\theta}{r}\right) \sin(\omega t - \beta r) \quad (f \text{ in Mc}) \quad (12\text{-}75)$$

where the direction of $\underline{H}$ is indicated by the unit vector $\underline{\varphi}_1$, whose direction at each and every point is perpendicular to the plane containing r (= OQ) and Z.

The energy always flowing away from the antenna which crosses any unit area perpendicular to $\underline{r}$ in unit time is

$$\underline{P} = \underline{E} \times \underline{H} = \underline{r}_1 \left(0.16 \, \frac{A^2 f^2 \sin^2\theta}{r^2}\right) \sin^2(\omega t - \beta r) \quad (12\text{-}76)$$

watts per square meter (f in Mc), where $\underline{r}_1$ indicates that the flow is in the $\underline{r}$ direction (away from the antenna).[20] At low frequencies this is negligible, and energy flow is composed of energy flowing alternately out and in, with no net radiation, as indicated in footnote 19.

20. Note that E and H are directly proportional to f, and P to f^2.

12-15. Energy Radiated by Short Antenna; Radiation Resistance.- Since there is no dissipation, absorption, or generation of energy in the assumed medium about the antenna, the total energy radiated per unit time can be obtained by integrating $\underline{P}$ as given by (12-76) over the surface of any sphere, radius r, center at the origin O. An element of area (zone about Z-axis) of such a sphere is

$$d\underline{S} = \underline{r}_1\, 2\pi r^2 \sin\theta \, d\theta$$

and the integral of $\underline{P} \cdot d\underline{S}$ for θ ranging from 0 to π gives the instantaneous total energy crossing the spherical surface. Inserting the average value of $\sin^2(\omega t - \beta r)$, which is equal to 1/2, yields the average energy radiated per unit time:

$$P_r = 0.69\, A^2 f^2 \quad \text{watts} \quad (f \text{ in Mc}) \qquad (12\text{-}77)$$

Thus for the short antenna, omitting time factors,

$$\left.\begin{aligned} H &= H_\varphi = \sqrt{P_r}\,\frac{0.025 \sin\theta}{r} && \text{amperes per meter} \\ E &= E_\theta = \sqrt{P_r}\,\frac{9.52 \sin\theta}{r} && \text{volts per meter} \\ P &= P_r\,\frac{0.24 \sin^2\theta}{r^2} && \text{watts per square meter} \\ P_r &= 0.69\, A^2 f^2 && \text{watts} \end{aligned}\right\} (12\text{-}78)$$

"Radiation resistance[21] is the quotient of the power radiated by an antenna by the square of the effective antenna current measured at the point where the power is supplied to the antenna." Thus for a short antenna the radiation resistance is

$$R_r = \frac{P_r}{I_{rms}^2} = 0.69 \left(\frac{A}{I_{rms}}\right)^2 f^2 = 1579\,\frac{h^2}{\lambda^2} \quad \text{ohms} \qquad (12\text{-}79)$$

where λ is wavelength and h has been written in place of ds for the length of the very short antenna (it will be recalled that A = I ds/4π and up to this point the various formulas developed have been for a very short antenna only).

12-16. Current Distribution Along a Linear Antenna.- The magnetic field $\underline{H}$ given by (12-75) is the field due to a current i in a length of wire $\underline{k}$ ds, provided A is replaced by i ds. Actual antennas being of finite height, we may either consider that (12-75) is an approximation to an actual antenna when A is replaced by Iℓ, where ℓ is the length of the antenna and I the current at some point in the antenna, or we can be more accurate

21. Standards on Transmitters and Antennas, I.R.E., 1938, p. 9. By "effective current" is meant the quantity often called the rms current, indicated in (12-79) by I_{rms}.

and consider the antenna to be composed of elements of length ds, find the field due to the current in each element, and sum (integrate) to find the field due to the entire length ℓ. The latter method requires that the current in each element ds of the antenna be known.

A thin straight-wire (linear) antenna is quite similar to a transmission line in so far as current distribution is concerned. Consider a single-wire transmission line parallel to the earth, which will be assumed a perfect conductor. If the load or receiving end of the line is open-circuited, and a sine-wave voltage applied at the sending end, it has been shown in Ch. 11 that there will be a distribution of current and voltage along the line which, when the attenuation constant α is small, will be a standing-wave distribution. This is illustrated in Ch. 11.

A linear antenna is a straight wire whose electrical length (i.e., length measured in wavelengths) is usually much greater than that of a 60-cps power transmission line. Whether the antenna is vertical, horizontal, or otherwise disposed, it has effective resistance, inductance, capacitance, and leakage per unit length which are substantially uniform. In addition, radiation resistance, which can be conceived as a fictitious circuit element accounting for the energy lost to the system by radiation, may likewise be considered to be distributed approximately uniformly along the antenna.

Experiment [22] shows that the net result of all these factors is to produce current and voltage distributions quite similar to those on transmission lines. The agreement is not exact, but is good enough for most practical purposes. The capacitance at the end of an antenna, resulting from the cessation of uniformity so that the end may be considered to form a capacitor with a large part of nearby conductors such as the earth, sometimes causes important differences, and the lack of constancy of parameter values per unit length may likewise give rise to discrepancies. These and other causes may result in a specific antenna having a length different from that calculated on a free-space wavelength basis, for example "half-wave" antenna may be 0.48 wavelength, or some other length, long. But, on the whole, a linear antenna excited by a sinusoidal voltage will have a distribution of current which approximates closely that which would exist on a transmission line under similar conditions of excitation, and termination.

In the following treatment, a standing-wave sinusoidal

22. Gihring, H.E., and Brown, G.H., General Considerations of Tower Antennas for Broadcast Use, Proc. I.R.E., 23, No. 4, April (1935). Numerous other papers on current distribution in antennas have been published.

distribution of current on a linear antenna will be assumed unless otherwise stated specifically.

12-17. Fields Due to a Linear Antenna of Finite Length.- In all the discussions of Sections 12-15 and 12-16 the factor A has appeared, and it will be recalled that if the instantaneous current i in a length ds of the linear antenna is varying sinusoidally ($i = I \cos \omega t$), then $A = Ids/4\pi$ and the fields (12-74) and (12-73) are the E and H radiation fields, respectively, due to the current in this element ds of the antenna. We can now list three aspects of the problem of the linear antenna of finite length.

1. If the length ℓ of the antenna is small in comparison with the wavelength λ, ds may be replaced by ℓ and the fields given by (12-73) and (12-74) will be reasonably good approximations.

2. If $\ell \ll \lambda$ [as in (1)] or if ℓ is not small compared with λ but is $< \lambda$, then ds may sometimes be replaced by an "effective height" h of the antenna. This approximation has been used particularly in connection with relatively low radio frequencies (e.g., frequencies less than those of the commercial broadcast band, order of magnitude of 1 Mc).

"The effective height of a grounded antenna is defined as $\int I\,ds$ divided by I_o, where I is the total vertical current in the antenna structure at height s, and I_o the current at the point where power is transferred to or from the connected apparatus."[23] Thus $I_o h = \int_0^\ell I\,ds$. Now if the antenna is vertical, and the field in the plane $\theta = 90^\circ$ (horizontal plane) is considered, then replacing A by $\sqrt{2}\, I_o h/4\pi$ in (12-78) yields

$$h = \frac{rE}{I_o \mu\omega} = \frac{rE}{1.26\, f I_o} \quad \text{meters} \quad (f \text{ in Mc}) \qquad (12\text{-}80)$$

where E is the rms field strength of the radiation field, r is distance in meters from the antenna to the point in the horizontal plane at which E is being measured, and I_o is the rms value of the sinusoidal input current to the antenna. If the antenna is grounded, or if the effect of the ground (earth) must be taken into account, or if the atmosphere does not act exactly as a perfect insulator but absorbs some energy, a factor may have to be introduced to allow for the change in E due to these causes. Aside from this factor, (12-80) indicates a method for determining h experimentally.

Use of the effective height h permits an allowance for the variation of current along the antenna, but makes no allowance

23. Standards on Transmitters and Antennas, I.R.E., 1938, p. 28, section 5.

for the fact that the distances from any point Q to various parts of the antenna are different except (practically) when Q is in a plane perpendicular to and passing through the antenna. These differences in "path length" are quite important in directional patterns, and consequently the use of effective height h is not a substitute for the more exact results to be given in (3) following.

3. A more comprehensive method of taking into account the finite length of a linear antenna is to determine the total field at any point by summing the individual contributions due to the current in each element of length ds of the antenna. This method includes (1) as a special case and gives a more exact result than (2). It is particularly important in obtaining the directional pattern of an antenna, to be discussed in the next section.

We will assume, in accordance with the discussion in Sec. 12-16, that the instantaneous current i in each element of length ds of the antenna varies sinusoidally with time ($i = I \cos \omega t$) and that I varies along the length ℓ of the antenna in such a manner that

$$I = I_m \sin \beta (\ell - s) \tag{12-81}$$

that is, the current in the antenna constitutes a standing wave. I_m is the value at a loop, and the input current to the antenna is $I_o \sin \beta(\ell - s_o)$ where s_o is the distance from the end of the antenna from which s is measured to the point on the antenna where the input line is connected.

Referring to Fig. 12-5, the distance r from an element ds of the antenna to any point Q is

$$r = R - s \cos \theta \tag{12-82}$$

and this is so closely equal to R that no appreciable error will be introduced in using R for r in most cases. But in $\sin (\omega t - \beta r)$ the total value of r is inconsequential, since any integer multiple of 2π may be subtracted from βr without changing the value of the sine. A relatively small change in r may produce a relatively large change[24] in $\sin (\omega t - \beta r)$. For this reason

24. For example, if $\lambda = 10$ meters, $\ell = 5$ meters, R = 3000 meters and r = 3003 meters, the change r - R is one part in 1000, but $\sin (\omega t - \beta r) = \sin (\omega t - 0.6\pi)$ and $\sin (\omega t - \beta R) = \sin \omega t$ so that the change in the sine is relatively great.

To look at the matter in a physical manner, consider radiation leaving an element at the bottom of the antenna (distance R to Q) and radiation leaving at the same time an element at s (distance r to Q). The former must travel a greater distance R-r, which at velocity c will require an additional time $(R-r)/c$. Thus the two radiations will be out of phase by a corresponding amount, and will arrive at Q at different times. The radiation from the element at 0 on the antenna arriving at Q at the same time as that from the element ds at s will have started $(R-r)/c$ earlier, and at this earlier time the current in the antenna was different in time phase (ωt) from what it was when the radiation at s started.

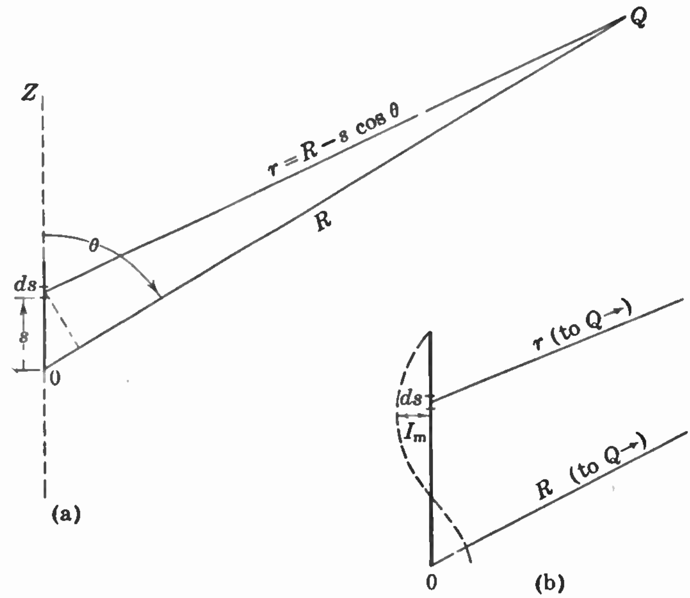


FIG. 12-5. (a) Antenna, length l, ds any element of length of the antenna at distance s from 0; Q is point at which radiation is to be calculated. (b) Enlarged view of antenna. showing assumed distribution (dashed) of *rms* value of current along antenna; I_m is the maximum value of the standing wave.

it is necessary to use (12-82) in sin ($\omega t - \beta r$) but not necessary elsewhere. By (12-66),

$$\text{H due to i ds} = \frac{I\ \beta}{r} \sin\theta \sin(\omega t - \beta r)\, ds \tag{12-83}$$

Hence, using (12-82),

$$\left.\begin{matrix}\text{H due to total cur-}\\ \text{rent in antenna}\end{matrix}\right\} = \frac{\beta \sin\theta}{R} \int_0^{\ell} \sin\beta(\ell - s)\sin(\beta R - \beta s \cos\theta - \omega t)\, ds$$

$$= \frac{\beta \sin\theta}{R} \int_0^{\ell} \sin(a - \beta s)\sin(w - bs)\, ds \tag{12-84}$$

where $a \equiv \beta\ell$, $w \equiv \omega t - \beta r$, and $b \equiv \beta \cos\theta$.

Since $2 \sin x \sin y = \cos(x-y) - \cos(x+y)$,

$$\frac{2HR}{\beta \sin\theta} = \int_0^{\ell} \cos[a - w - (\beta - b)s]ds - \int_0^{\ell} \cos[a + w - (\beta + b)s]ds$$

and since each integral is of the form of the integral of cos(m + ns) ds, it is a matter of integrating, substituting limits, and manipulating to show that

$$H R \sin\theta = (\cos b\ell - \cos\beta\ell)\sin w + (\cos\theta \sin\beta\ell - \sin b\ell)\sin w \qquad (12\text{-}85)$$

or finally
$$H = \frac{I_m}{R} F(\theta) \sin(w + \delta) \qquad (12\text{-}86)$$

where $F(\theta)$ is the directional factor (gives the variation of magnitude of H with θ):

$$F(\theta) = \frac{1}{\sin\theta}[1 + \cos^2\beta\ell - 2\cos b\ell \cos\beta\ell - 2\cos\theta \sin b\ell \sin\beta\ell + \cos^2\theta \sin^2\beta\ell]^{\frac{1}{2}} \qquad (12\text{-}87)$$

Values of $F(\theta)$ for specific cases are given in the following Table 12-V and curves of $F(\theta)$ are shown in Fig. 12-6.

TABLE 12-V

Directional Factors of Isolated Linear Antenna

Antenna	ℓ	$\beta\ell$	$F(\theta)$	Eq.
Half-wave	$\frac{1}{2}\lambda$	π	$\frac{2\cos(\frac{\pi}{2}\cos\theta)}{\sin\theta}$	12-87a
Very short	$\ell << \lambda$		$\beta\ell \sin\theta$	12-87b
Any odd number of half-wavelengths	$(2n+1)\frac{\lambda}{2}$	$(2n+1)\pi$	$\frac{2\cos(\frac{n\pi}{2}\cos\theta)}{\sin\theta}$	12-87c
Any even number of half-wavelengths	$n\lambda$	$2n\pi$	$\frac{2\sin(\frac{n\pi}{2}\cos\theta)}{\sin\theta}$	12-87d
Quarter-wave	$\frac{1}{4}\lambda$	$\frac{\pi}{2}$	$\frac{\cos^2\theta - 2\cos\theta\sin(\frac{\pi}{2}\cos\theta)}{\sin\theta}$	12-87e
Three quarter-wave-lengths	$\frac{3}{4}\lambda$	$\frac{3\pi}{2}$	$\frac{\cos^2\theta + 2\cos\theta\sin(\frac{3\pi}{2}\cos\theta)}{\sin\theta}$	12-87f
Any length:	$F(\theta) = \frac{1}{\sin\theta}[1+\cos^2\beta\ell+\cos^2\theta\sin^2\beta\ell - 2\cos b\ell\cos\beta\ell - 2\cos\theta\sin b\ell\sin\beta\ell]^{\frac{1}{2}}$ where $b \equiv \beta\cos\theta$			12-87g

These curves are directional patterns. "The directional pattern of a transmitting antenna is the polar characteristic which indicates the intensity of the radiation field at a fixed distance in different directions in space[25]....." The most general information which could be given would be a three-dimensional diagram showing radiation in every direction in space.[26] It is

25. Standards on Transmitters and Antennas, I.R.E., 1938, p. 30. The paragraph from which the quotation is made continues, "Similarly, the directional pattern of a receiving antenna is the polar characteristic which indicates the response of the antenna to unit field intensity from different directions. For a given antenna the two characteristics are identical in shape."

26. Examples of three-dimensional radiation directional patterns are given in Foster, Ronald M., Directive Diagrams for Antenna Arrays, B.S.T.S., 5, No. 2, April (1926), and in Southworth, G.C., Certain Factors Affecting the Gain of Directive Antennas, Proc. I.R.E., 18, No. 9, Sept. (1930).

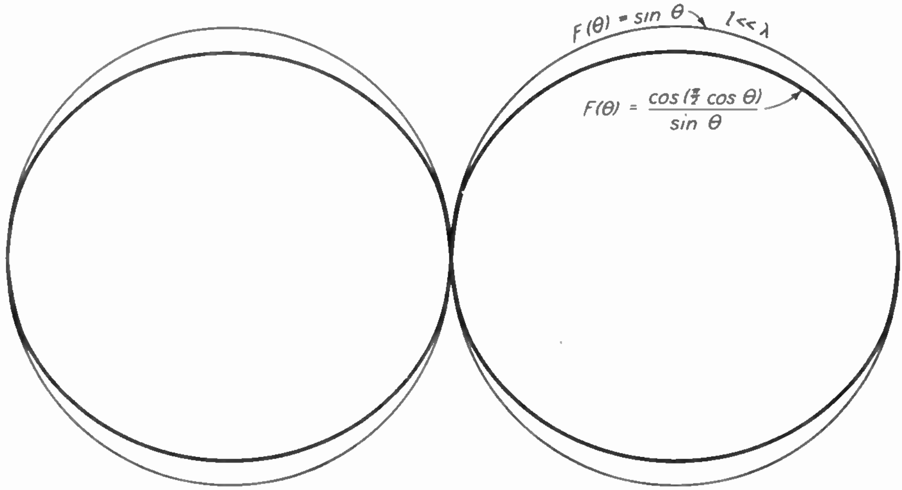


FIG. 12-6a. $\ell << \lambda$ (outer curve)
$\ell = \frac{1}{2}\lambda$ (inner curve)

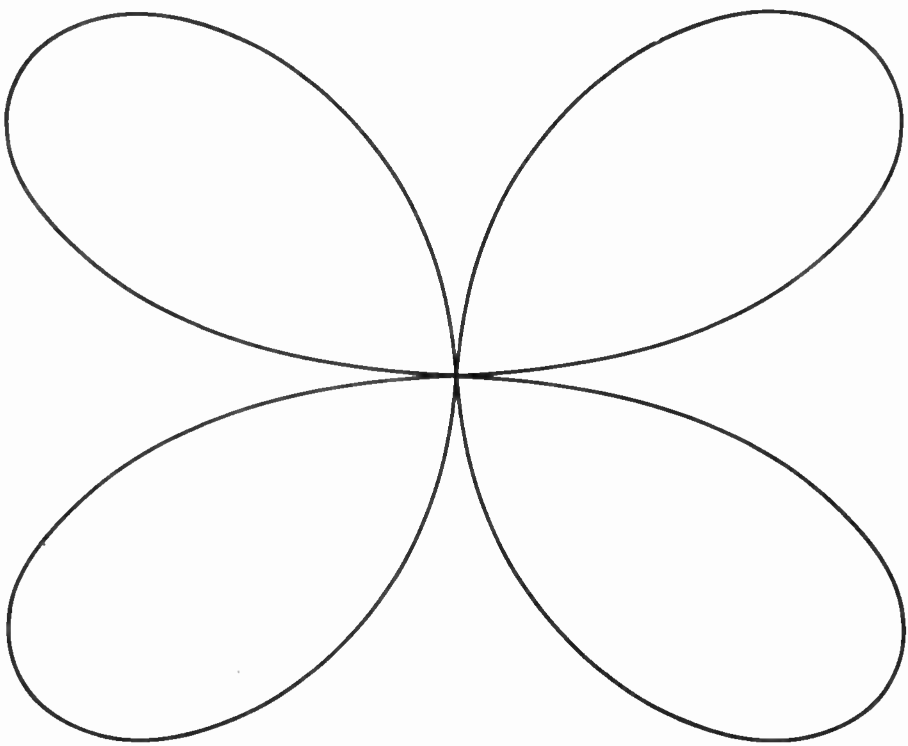


FIG. 12-6b. $\ell = \lambda$

$$F(\theta) = \frac{\sin(\pi \cos \theta)}{\sin \theta}$$

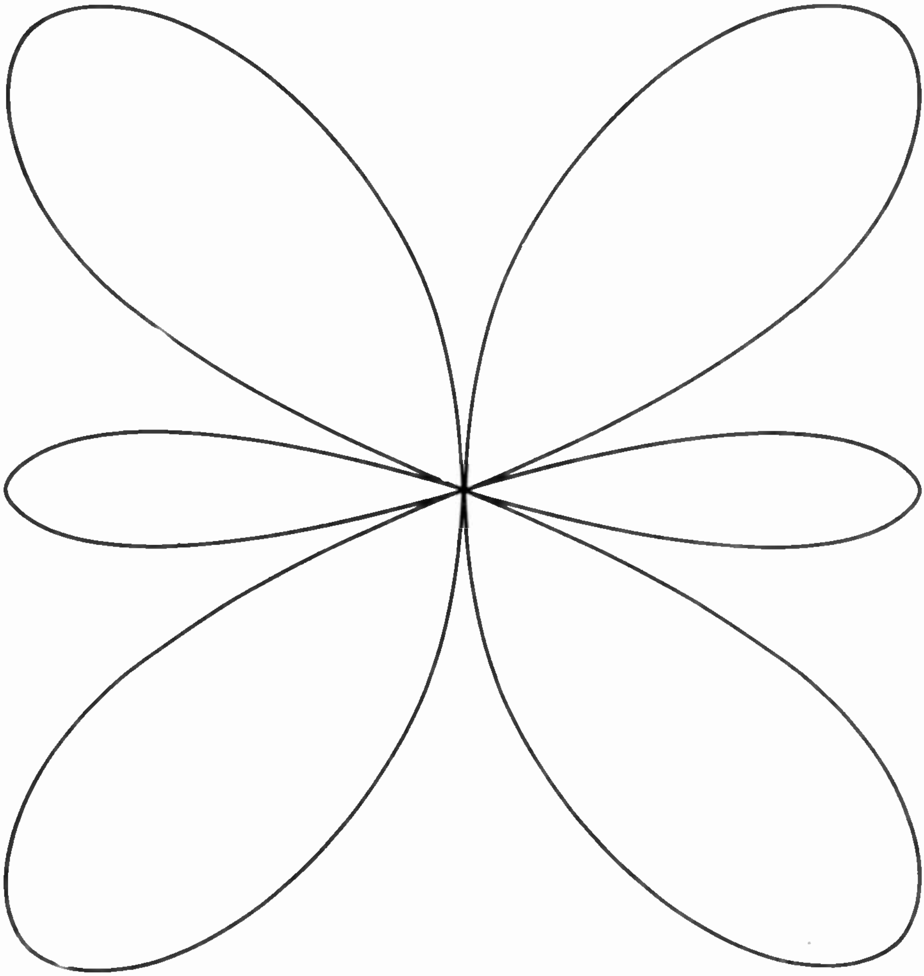


FIG. 12-6c. $l = \frac{3}{2}\lambda$

$$F(\theta) = \frac{\cos\left(\frac{3\pi}{2}\cos\theta\right)}{\sin\theta}$$

seldom feasible to do this, and in consequence, two-dimensional patterns showing radiation in a given plane are used. Although Fig. 12-6 is labeled "Vertical Radiation Patterns, etc." it is well to bear in mind that this is conventional jargon. The antenna can be placed in any position, and the patterns shown are those which obtain in any plane containing the antenna. The "horizontal pattern" of a vertical antenna is simply a circle, from symmetry. But if the antenna is placed in a horizontal position, the directive pattern that is called the vertical

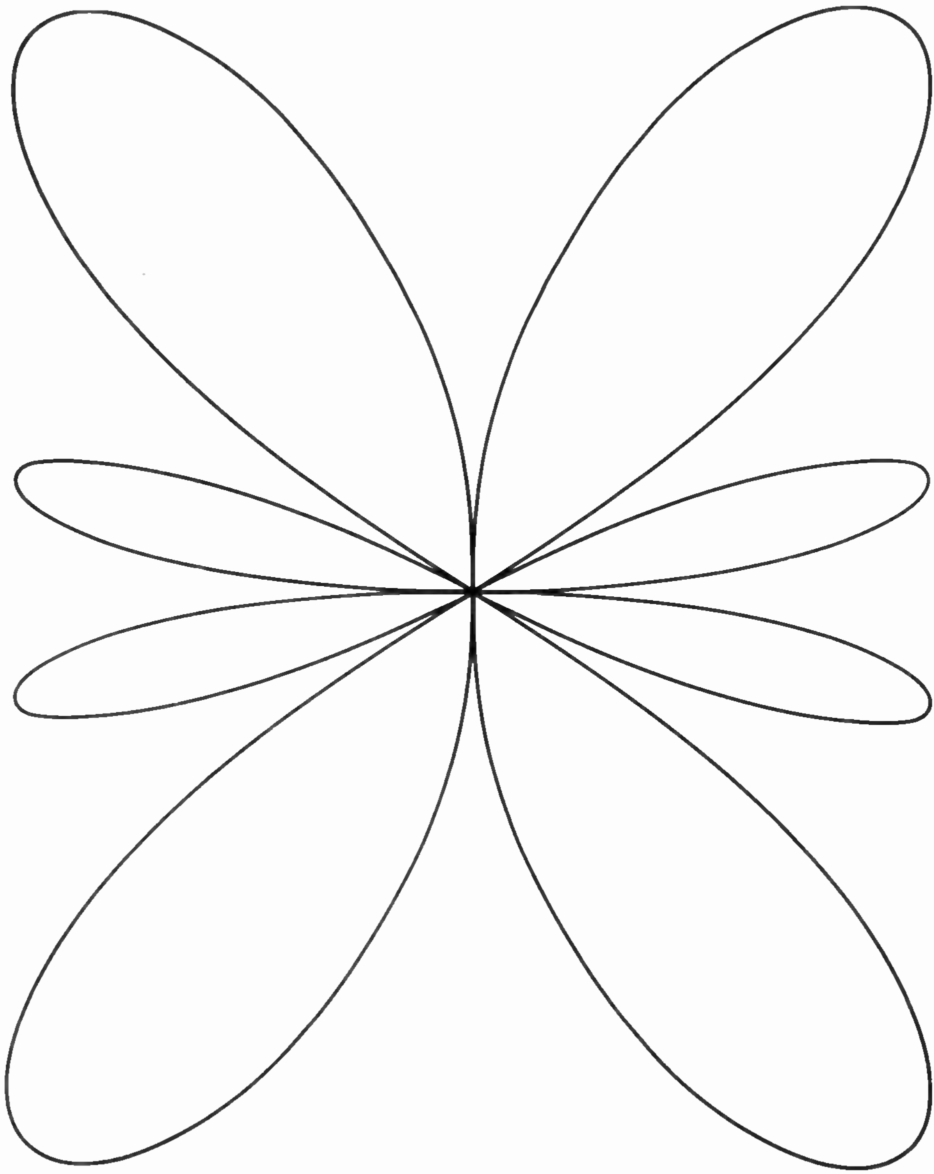


FIG. 12-6d. $\ell = 2\lambda$

$$F(\theta) = \frac{\sin(2\pi \cos\theta)}{\sin\theta}$$

radiation pattern in Fig. 12-6 becomes the actual horizontal pattern, and it is seen that a half-wave horizontal linear antenna has a distinct directional pattern in the horizontal plane whereas a vertical half-wave antenna radiates equally well in all directions in the horizontal plane.

12-18. Effect of Ground or of a Perfect, Flat Sheet, Reflector.- In many cases--for example, that of commercial broadcast antennas--the ground can often be assumed to be a perfect conductor. At ultra-high frequencies, flat sheets of conducting material are sometimes used in back of an antenna as reflectors. The problem of determining radiation at a given point when the antenna is attached to or near a perfectly conducting flat sheet may be handled much like that of the radiation from an "isolated" linear antenna (Sec. 12-17) by recalling two propositions of electromagnetic theory:

1. At the bounding surface of a perfect conductor the component of $\underline{E}$ tangent to the surface must be zero. The component of $\underline{E}$ normal to the surface that results when a wave strikes the surface is double what it would be if the surface were absent.

2. Any arrangement of conductors, surfaces, sources, etc., which produces the same conditions over the surface or surfaces of a given region as are produced by a different arrangement of surfaces, sources, etc., will produce the same fields within that region.

Figure 12-7b shows how a vertical antenna, connected to the ground at its base, and an "image antenna" with a current distribution similar to the first, but opposite in phase, will produce above ground fields exactly equal to those produced by the actual antenna attached to the perfectly conducting earth (Fig. 12-7a). It is not necessary that the antenna be grounded, and the argument holds equally well for antennas above the earth, and for antennas in front of large flat sheets of conducting material (perfect conductivity assumed).

Considering the actual problem posed by Fig. 12-7a, and using the equivalent shown in Fig. 12-7c, the magnetic field H at Q is the sum of the fields due to the actual antenna and to the image:

$$H = \beta\int_0^{\ell} \frac{I \sin\theta}{r} \sin(\omega t - \beta r)ds + \beta\int_{-\ell}^{0} \frac{I \sin\theta}{r} \sin(\omega t - \beta r)ds$$

$$= \frac{\beta I_m \sin\theta}{R}\left\{\int_0^{\ell} \sin\beta(\ell-s)\sin[\beta(R-s\cos\theta)-\omega t]\,ds\right.$$

$$\left.+\int_{\ell}^{0} \sin\beta(\ell+s)\sin[\beta(R-s\cos\theta)-\omega t]\,ds\right\} \qquad (12\text{-}88)$$

which, after some reduction, yields

$$H = \frac{2I_m}{R}\,\frac{\cos(\beta\ell\cos\theta)-\cos\beta\ell}{\sin\theta}\sin(\omega t-\beta r) \qquad (12\text{-}89)$$

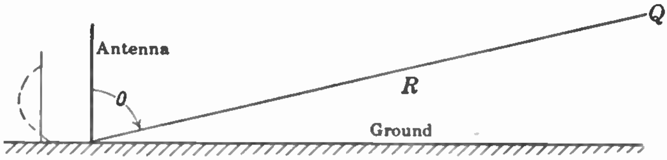
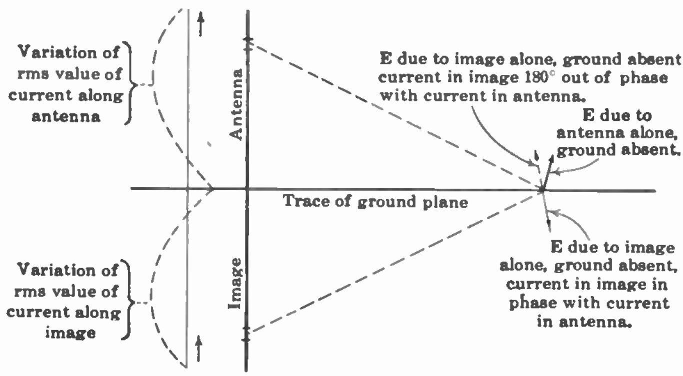


FIG. 12-7a. The original problem: Vertical antenna connected to ground (assumed perfect conductor); current distribution in antenna shown at left; fields at Q to be determined.



Small arrows indicate that the instantaneous current in the image is 180° out of phase, relatively, with the instantaneous current in the antenna.

Thus for the antenna, $i = I_m \sin \beta(l-s) \cos \omega t$,

and for the image $i = I_m \sin \beta(l+s) \cos \omega t$,

i for the antenna being assumed positive in the direction of the arrow, and i for the image being assumed positive in the direction of the arrow on it, which is not the true image direction but opposite to it.

FIG. 12-7b. Illustration of how antenna and image produce the same conditions over a surface as exist when that surface is the surface of a perfect conductor.

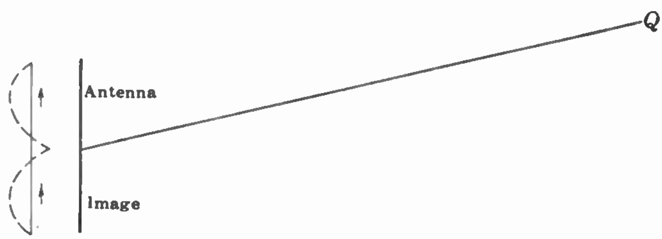


FIG. 12-7c. The equivalent problem, solution of which yields the solution of the problem of Fig. 12-7a.

or

$$H = \frac{2I_m}{R} F(\theta) \sin(\omega t - \beta r) \qquad (12\text{-}90)$$

where

$$F(\theta) = \frac{\cos(\beta l \cos\theta) - \cos\beta l}{\sin\theta}$$

gives the directional pattern of the antenna. If we plot $F(\theta)$ against θ (usually on a polar diagram, as previously indicated, but in u-h-f work it is sometimes preferable to plot in a rectangular system, using θ for abscissa), the relative radiation at points a given distance R from the antenna is obtained. Figure 12-8

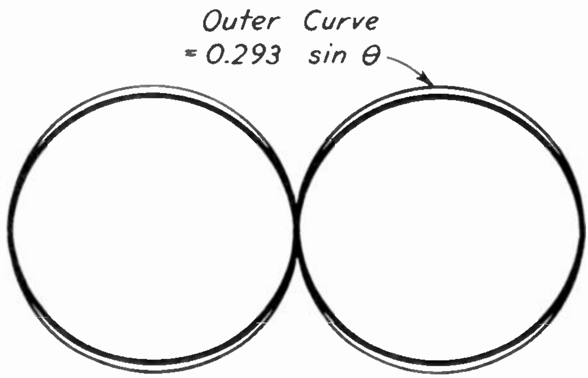


FIG. 12-8a. $l = \frac{1}{8}\lambda$

$$F(\theta) = \frac{\cos(\frac{\pi}{4}\cos\theta) - 0.707}{\sin\theta}$$

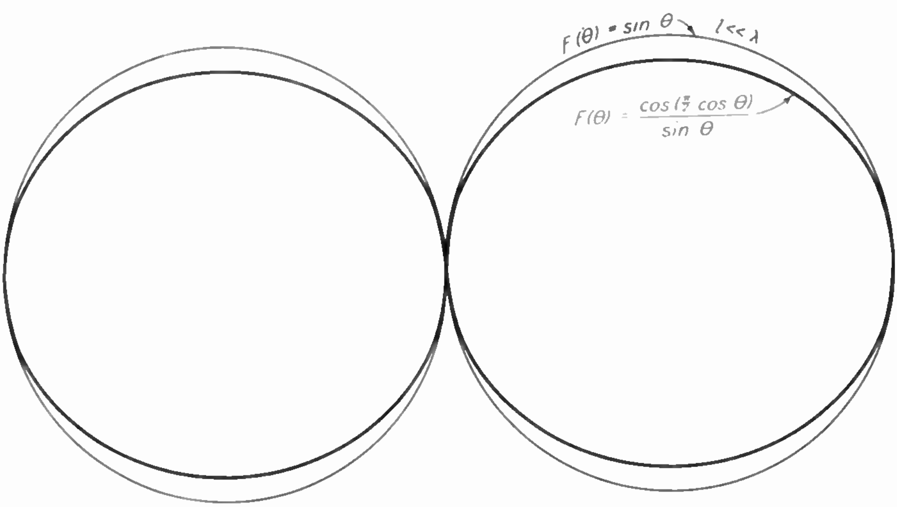


FIG. 12-8b. $l = \frac{1}{4}\lambda$

$$F(\theta) = \frac{\cos(\frac{\pi}{2}\cos\theta)}{\sin\theta}$$

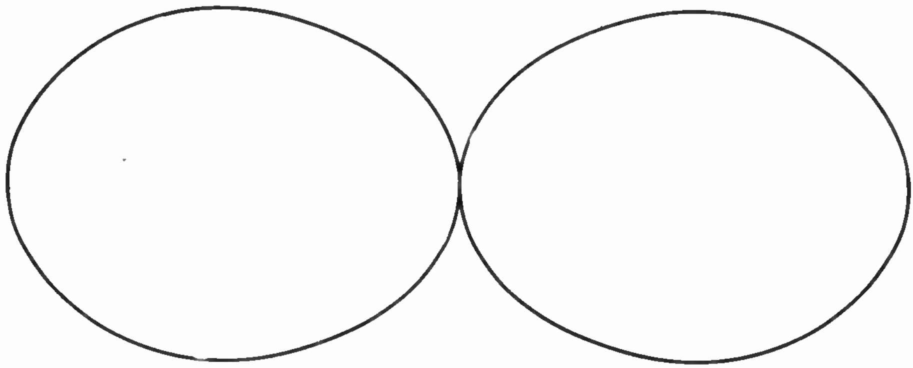


FIG. 12-8c. $\ell = \frac{3}{8}\lambda$

$$F(\theta) = \frac{\cos(\frac{3}{4}\pi \cos\theta) + 0.707}{\sin\theta}$$

[$F(\theta)/1.707$ plotted]

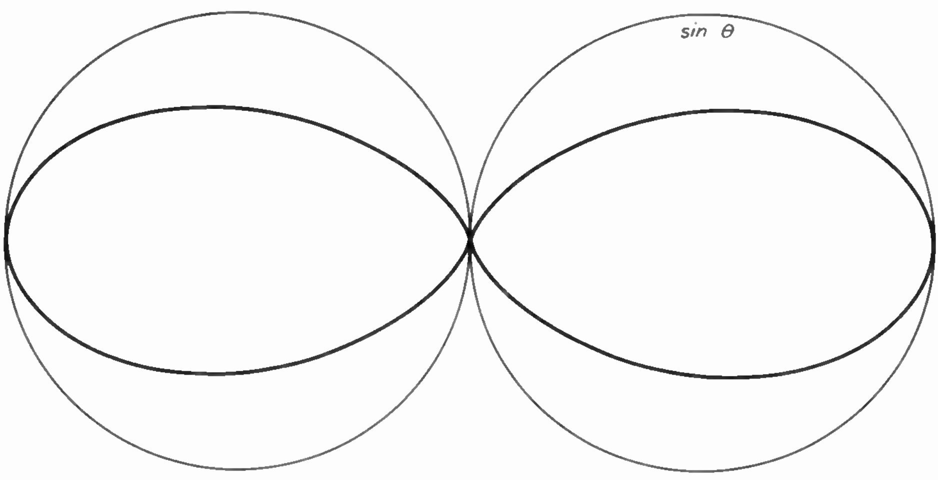


FIG. 12-8d. $\ell = \frac{1}{2}\lambda$

$$F(\theta) = \frac{\cos(\pi \cos\theta) + 1}{\sin\theta}$$

[$F(\theta)/2$ plotted]

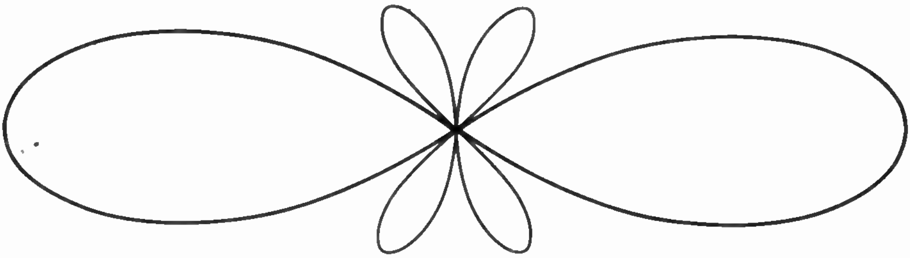


FIG. 12-8e. $\ell = \frac{5}{8}\lambda$

$$F(\theta) = \frac{\cos(\frac{5\pi}{4}\cos\theta) + 0.707}{\sin\theta}$$

[$F(\theta)/1.707$ plotted]

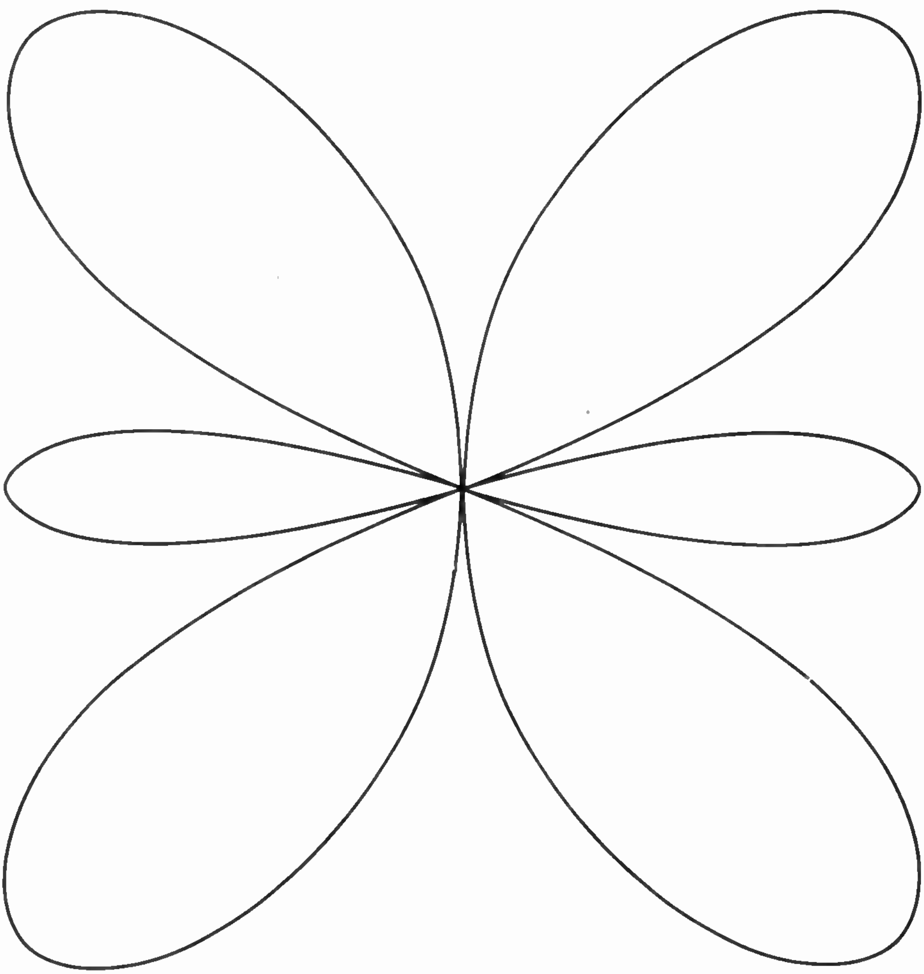


FIG. 12-8f. $\ell = \frac{3}{4}\lambda$

$$F(\theta) = \frac{\cos(\frac{3}{2}\pi\cos\theta)}{\sin\theta}$$

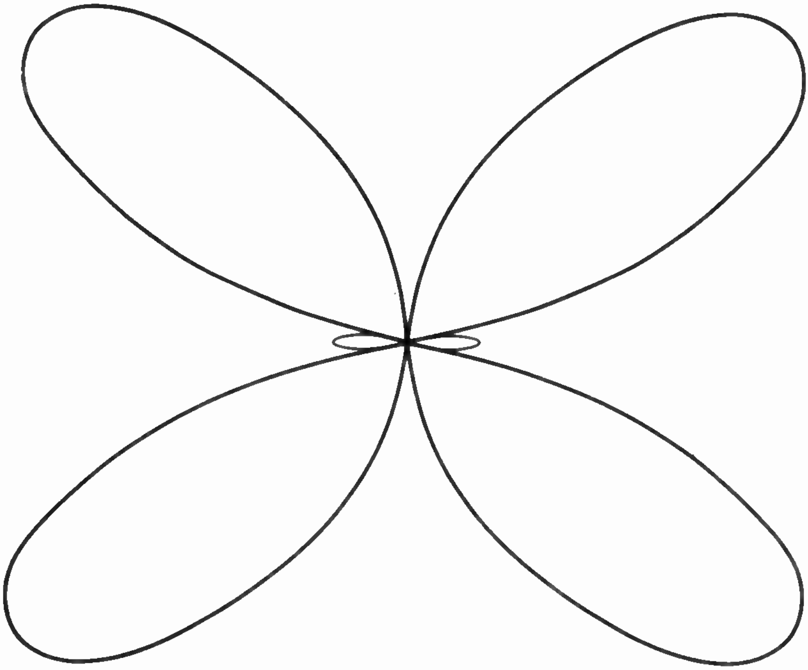


FIG. 12-8g. $\ell = \frac{7}{8}\lambda$

$$F(\theta) = \frac{\cos(\frac{7\pi}{4}\cos\theta) - 0.707}{\sin\theta}$$

[$F(\theta)/2$ plotted]

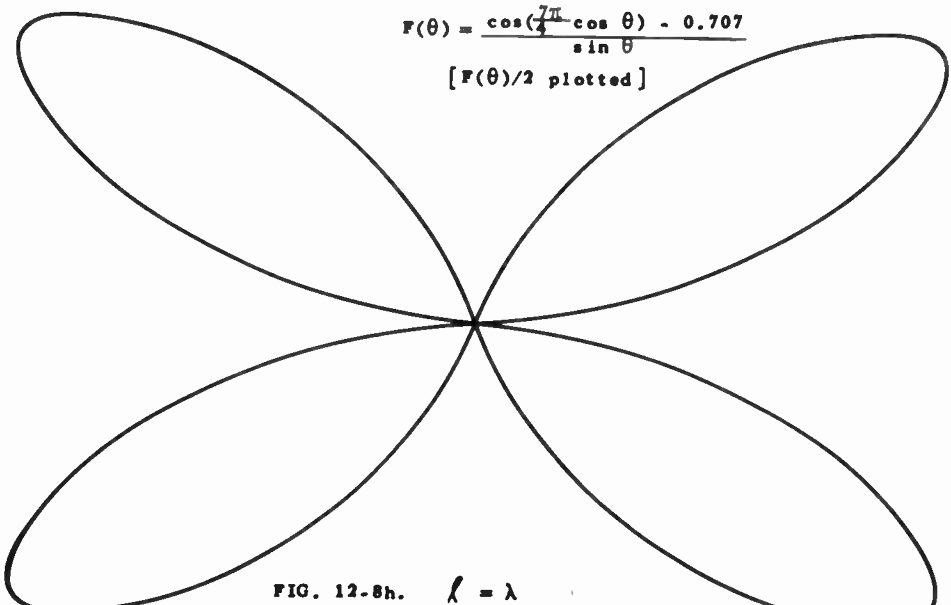


FIG. 12-8h. $\ell = \lambda$

$$F(\theta) = \frac{\cos(2\pi\cos\theta) - 1}{\sin\theta}$$

[$F(\theta)/2$ plotted]

shows a number of such directional patterns.

In the case of an antenna in front of a flat conducting sheet (of large extent), the patterns are halves of those for 180° phase difference given in Fig. 12-10.

12-19. Input Impedance of an Antenna.- At the terminals at which an antenna is to be fed, there is a definite impedance which loads the transmission line leading to the antenna. This impedance is the impedance of the antenna, and is a somewhat complicated function of antenna length.[27] Resistance is chiefly the radiation resistance, but all absorption of energy from the radiator contributes to the effective input resistance.

For an isolated linear oscillator of length $\ell = \lambda/2$, the radiation resistance is 72.5 ohms, and provided m is an integer > 2, it is given by

$$72.5 + 30 \log m \qquad (\ell = m\lambda/2) \tag{12-91}$$

Figure 12-9 shows the variation with frequency of particular broadcast antenna, and is introduced here primarily to emphasize this variation. For commercial broadcasting a band width of the order of 10 kc may be radiated, so that an antenna system matched at the carrier frequency is for practical purposes matched throughout the entire range of frequencies used. At ultra-high frequencies a band width of several megacycles may be required, and special measures must be taken so to broaden the impedance vs frequency characteristics of the antenna system that reasonably good matching can be obtained over the entire range of frequencies radiated. See footnotes 27 and 33.

12-20. Antenna Arrays.- It is convenient to distinguish the isolated linear antenna from all other antenna systems, although such a distinction is not always sharp. In the case of an isolated linear antenna, the chief characteristics under control are:

---

27. Resistance and reactance components of the impedance of a linear vertical antenna are given by Brown, G.H., and King, Ronald, High-Frequency Models in Antenna Investigations, Proc. I.R.E., 22, No. 4, April (1934), and by Brown, George H., A Consideration of the Radio-Frequency Voltages of Broadcast Tower Antennas, Proc. I.R.E. 27, No. 9, Sept. (1939). Stratton, J.A., "Electromagnetic Theory," McGraw-Hill Book Co., 1941, gives a concise derivation of the radiation resistance formula. The reader will recognize that the formula for radiation resistance (12-79) given earlier applies only to a very short antenna. Brown, in the 1939 paper, shows that large radius of antenna wire makes the tuning broader.

Kandoian, A.G., Radiating Systems and Wave Propagation, Electronics, 15, No. 4, April (1942), shows the variation of radiation resistance with antenna length (measured in wavelengths) up to 7.5 λ.

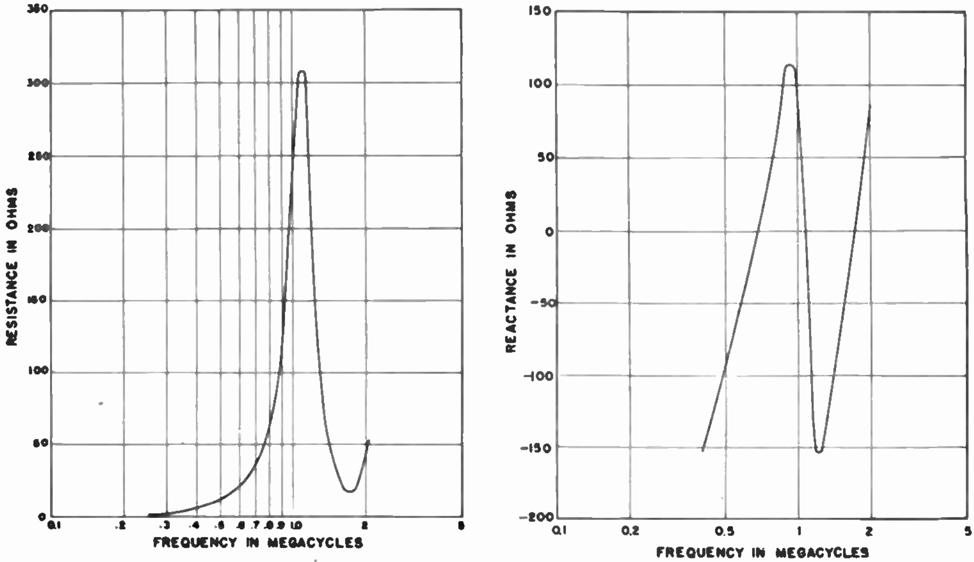


FIG. 12-9. Resistance and reactance characteristics of an antenna for use at a broadcast frequency (General Radio Experimenter, Feb. 1938).

1. Length of the antenna.

2. Current distribution, which can be changed by inserting inductors or capacitors in the antenna or at its end, by "folding" part of the antenna, by terminating the antenna in its characteristic impedance so that standing waves are not produced, etc.

3. Shape of the antenna wire (diameter, etc.), this is of particular importance in connection with the impedance at the input terminals of the antenna.

In antenna systems, involving two or more antennas, or a single antenna which is not linear, or a linear antenna with a reflecting system of sheet metal or its equivalent, or in general any arrangement except the isolated linear antenna, the following may be controllable:

1. Lengths of the antennas.

2. Relative positions of the antennas in simple arrays, or

3. Shape of the antenna structure (V, turnstile, loop, rhombic, etc.)

4. Relative phases of the currents in the antennas, or

5. Type of current distribution in the antennas.

6. Relative magnitudes of the currents in the antennas.

7. Positions and shapes of reflecting surfaces, if any.

As a consequence of the great leeway which the factors listed above allow, the statement is sometimes made that virtually any specified radiation pattern can be obtained, but in practice it is often a large step from the specification of a pattern to its realization within practical limits.

To introduce the reader to the directional patterns for some simple arrays, the formula for two isolated, similar, parallel linear antennas, perpendicular to a common plane which passes through their bases, will be derived. Consider (Fig. 12-10a) such a case, noting particularly in the figure the angle φ. The horizontal pattern is no longer a circle, as in the case of the single isolated antenna, and horizontal patterns now become important. Using the notation indicated on the figure, assuming standing waves of current, with I_m the maximum value of the standing wave distribution and ℓ the length of the antennas,

$$\left.\begin{array}{l}\text{H produced at Q by}\\ \text{the first antenna}\end{array}\right\} = \frac{I_m}{R} F_1(\theta) \sin(\omega t - \beta R + \delta)$$

where $F_1(\theta)$ is the $F(\theta)$ given by (12-87).

$$\left.\begin{array}{l}\text{H produced at Q by}\\ \text{the second antenna}\end{array}\right\} = \frac{I_m}{R} F_1(\theta) \sin(\omega t - \beta R + \beta d \cos\varphi \sin\theta + \alpha)$$

Since the magnetic fields at Q are in the same direction for practical purposes, the two may be added, giving for the total H at Q

$$H = 2\frac{I_m}{R} F_1(\theta) \cos\frac{1}{2}(\beta d \cos\varphi \sin\theta - \alpha) \sin(\omega t - \beta R + \delta') \quad (12\text{-}92)$$

or in magnitude

$$H = 2\frac{I_m}{R} F_1(\theta) \cos\frac{1}{2}(\beta d \cos\varphi \sin\theta - \alpha) = 2\frac{I_m}{R} F_2(\theta, \varphi) \quad (12\text{-}93)$$

where $F_2(\theta,\varphi) = F_1(\theta) \cos\frac{1}{2}(\beta d \cos\varphi \sin\theta - \alpha)$

For the half-wave case ($\ell = \lambda/2$, $\beta\ell = \pi$),

$$F_2(\theta,\varphi) = \frac{2\cos(\frac{\pi}{2}\cos\theta)}{\sin\theta} \cos\frac{1}{2}(\beta d \cos\varphi \sin\theta - \alpha) \quad (12\text{-}94)$$

and for a horizontal pattern ($\theta = 90^o$),

$$F_2(\varphi) = 2\cos(\frac{\pi d}{\lambda}\cos\varphi - \frac{1}{2}\alpha) \quad (12\text{-}95)$$

In Fig. 12-10b are shown plots of

$$\cos(\frac{\pi d}{\lambda}\cos\varphi + \frac{1}{2}\alpha) \quad (12\text{-}96)$$

for numerous values of d and α. More extensive tables will be found in papers by Campbell,[28] Foster,[29] and particularly by

28. Campbell, George A., U.S. patent specification, Sept. 1919, reprinted in Collected Papers of George A. Campbell, A.T. and T. Co., 1937.

29. Foster, Donald M., Directive Diagrams of Antenna Arrays, Bell. Sys. Tech. Jour., 5, No. 2, April (1926).

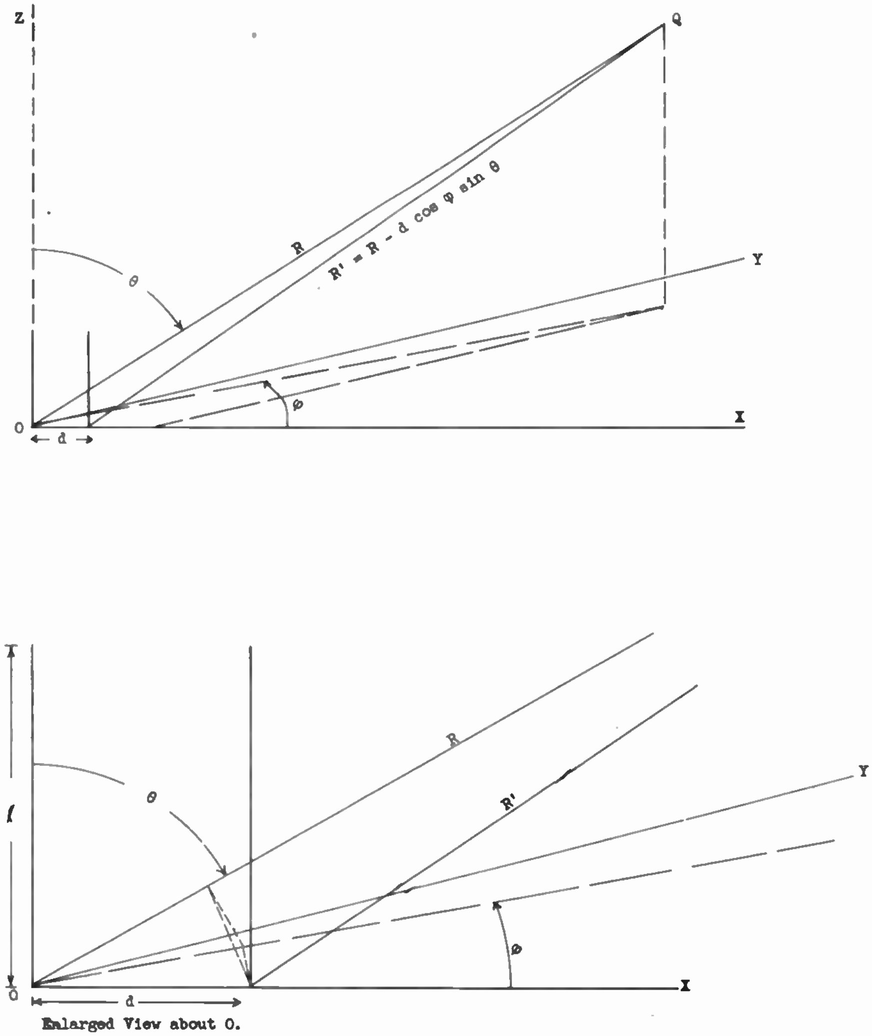


Fig. 12-10a.

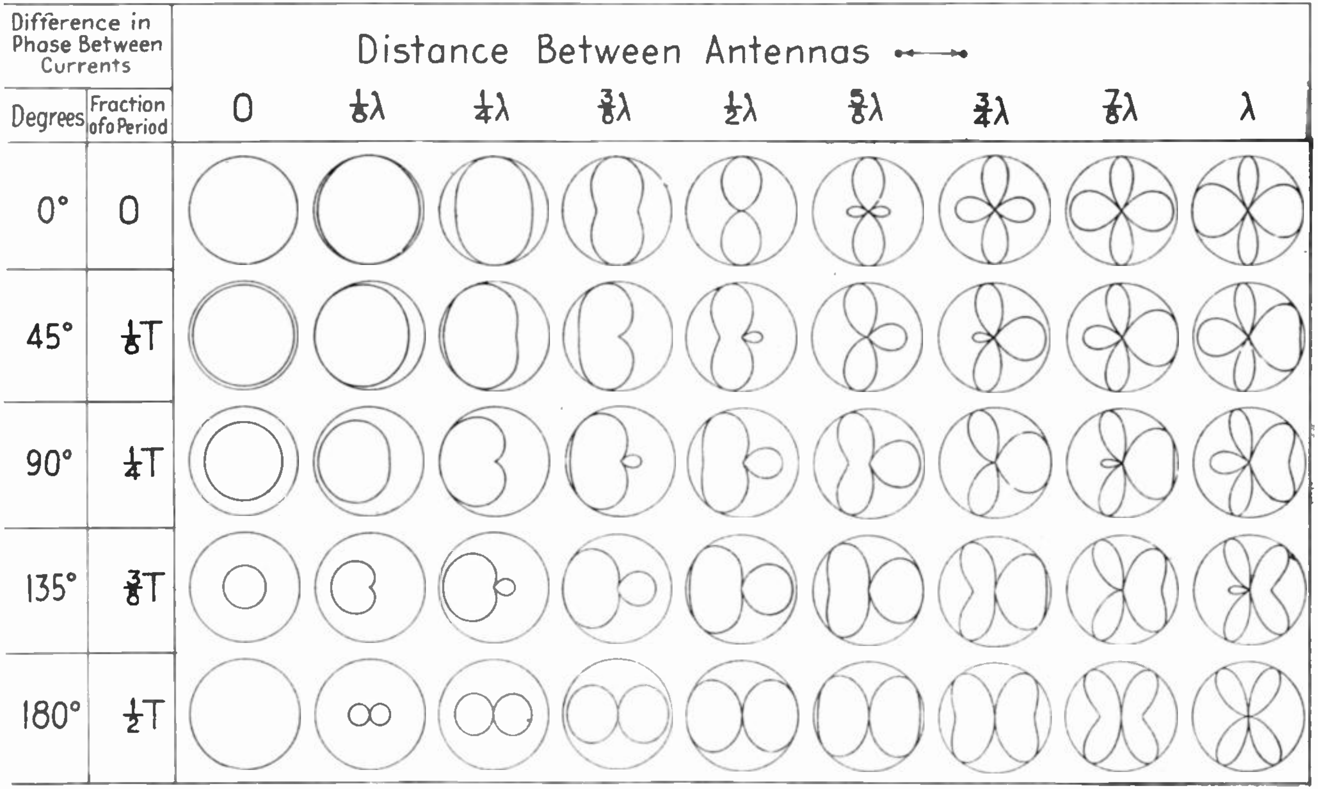


FIG. 12-10b. Horizontal patterns of two vertical half-wave antennas. Phase difference between currents in the antennas is indicated at left (current of right-hand antenna lags that of left-hand antenna for sending array); distance between antennas is indicated at top. (From Ronald M. Foster, B.S.T.J., April 1926.) These patterns do not take into account mutual coupling; for similar patterns making allowance under certain circumstances, for the effects of the antennas on one another, see G.H. Brown, Proc. I.R.E., Jan. 1937.

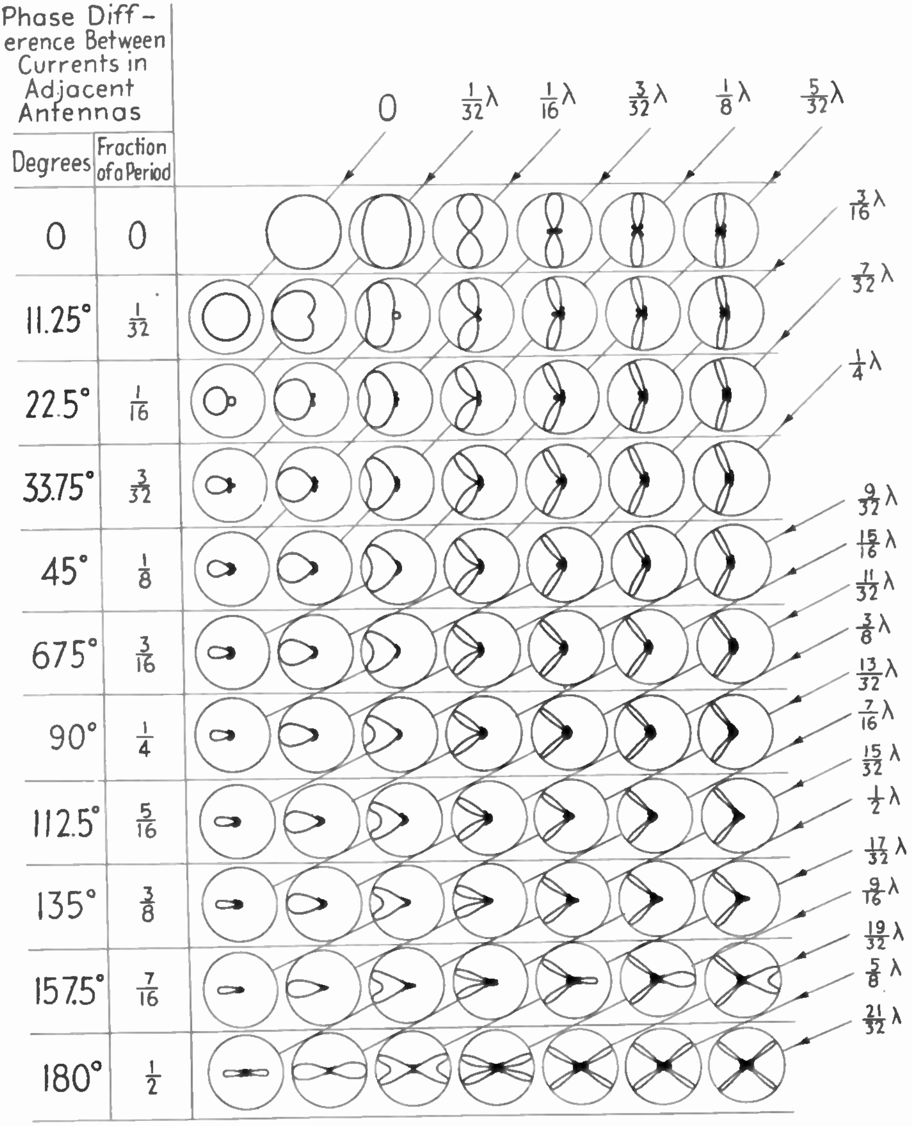


FIG. 12-11. Horizontal patterns, 16 vertical half-wave antennas, centers on a horizontal straight line; phase difference between currents in adjacent antennas shown on left, spacing between adjacent antennas indicated along diagonals. The circle accompanying each pattern is the radiation pattern of the 16 antennas when coincident with one another, and with all currents in phase, as indicated in the first position in the top row. (Adapted from a much more extensive table given by Ronald M. Foster, B.S.T.J. April 1926.)

Southworth[30] whose paper contains an extensive bibliography.

In the more general case of n half-wave antennas, parallel to one another, centers on a straight line, uniformly spaced $a(= d/\lambda)$ wavelengths apart and with currents differing in phase by α°, the radiation pattern in a plane perpendicular to the antennas and passing through their centers is[28]

$$F_n(\varphi) = \frac{\sin n(\pi a \cos \varphi + \alpha/2)}{n \sin(\pi a \cos \varphi + \alpha/2)} \qquad (12\text{-}97)$$

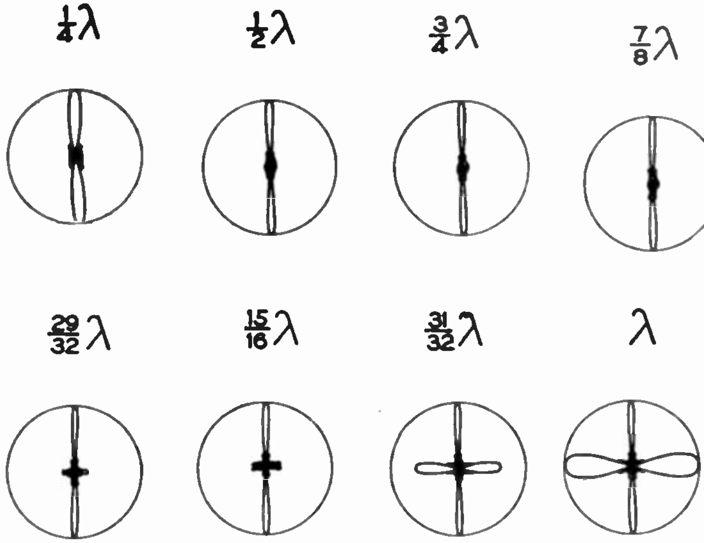


FIG. 12-12. Extension of the top row of Fig. 12-11 to show change in pattern for spacing up to one wavelength between adjacent antennas in the 16-element array; currents in phase. (From Foster, loc. cit.)

In Figs. 12-11 and 12-12 are shown some patterns for a 16-element array of this type.

Figures 12-13 and 12-14 show the effect of changing the number of antennas without changing appreciably the space occupied by the antennas. In Fig. 12-13 the phases are kept constant; in Fig. 12-14 the phases are changed in the same ratio as the distance between antennas is changed. In this figure it should be particularly noted that the pattern becomes more or less unidirectional, an important characteristic when it is

30. Southworth, G.C., Certain Factors Affecting the Gain of Directive Antennas, Proc. I.R.E., 18, No. 9, Sept. (1930).

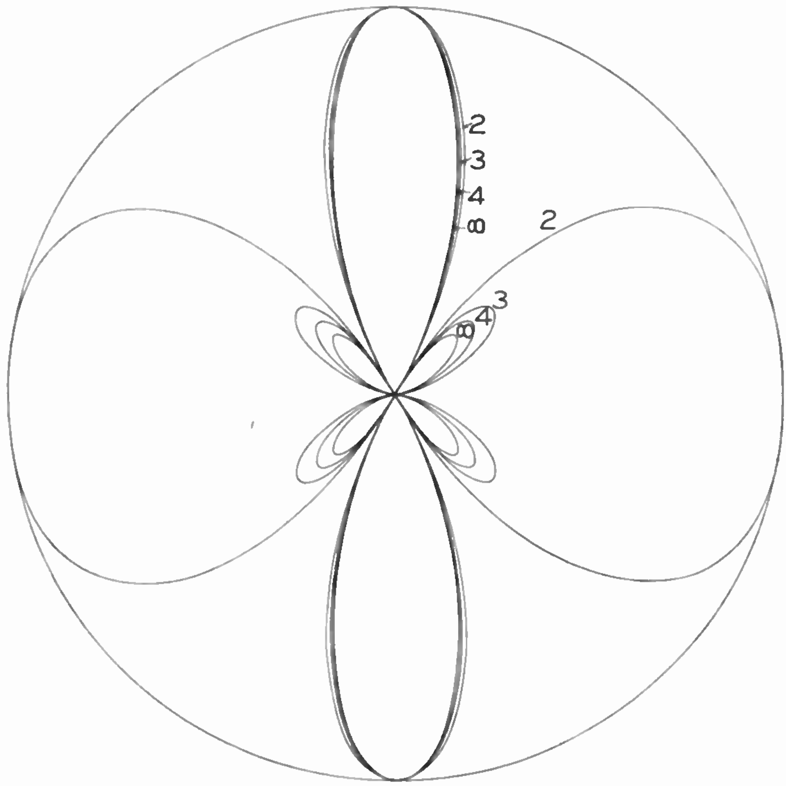


FIG. 12-13. Horizontal patterns for n vertical half-wave antennas, spaced $2\lambda/n$ apart along a horizontal line; currents in phase. The figure shows the effect of increasing the number of antennas used. The curve for $n = 2$ appears in Fig. 12-10 and the curve for $n = \infty$ is substantially the same as the curve for $n = 16$ which appears in Fig. 12-11. (From Foster, loc. cit.)

desired to send as much energy as possible to one point only. Southworth[30] gives a set of patterns, reproduced in Fig. 12-15, for two rows of parallel half-wave antennas, distance between rows $\lambda/4$, for which the directional pattern in a plane perpendicular to and bisecting the antennas is[30]

$$F(\varphi) = \frac{\sin(n\pi a \sin\varphi)}{n \sin(\pi a \sin\varphi)} \cos\left(\frac{\pi}{4}\cos\varphi - 1\right) \qquad (12\text{-}97a)$$

It will be noticed that the direction of transmission is "broadside" to the array. When the major transmission in the bisecting plane is in the direction of a row, the transmission is "endfire." Figure 12-11 shows both broadside and endfire results, as well as some not distinctly either.

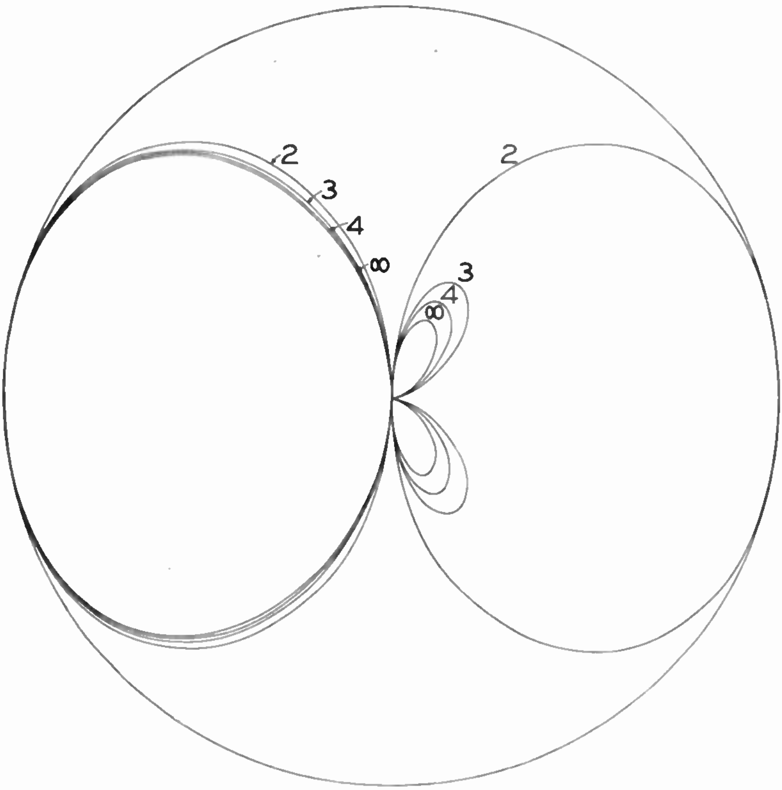


FIG. 12-14. Effect of changing spacing and phase difference proportionately. Horizontal patterns of n vertical half-wave antennas, spaced λ/n apart, currents in adjacent antennas $360°/n$ apart. The curve for $n = 2$ appears in Fig. 12-10 and that for $n = \infty$ is for practical purposes the same as that for $n = 16$, given in Fig. 12-11. (From Foster, loc. cit.)

12-21. Parasitic Antennas.- An antenna which is not driven, but receives its current by induction from one or more other antennas, is called a parasitic antenna. If, in Fig. 12-10a, a driving voltage is applied to one antenna and not to the other, the latter will have induced in it voltage and current resulting from the field--both local and radiation components--of the first antenna. The second antenna radiates, and the resultant field at distant points is that due to the radiation from both antennas.

Essentially such a pair constitutes a simple coupled circuit, and it is customary to write

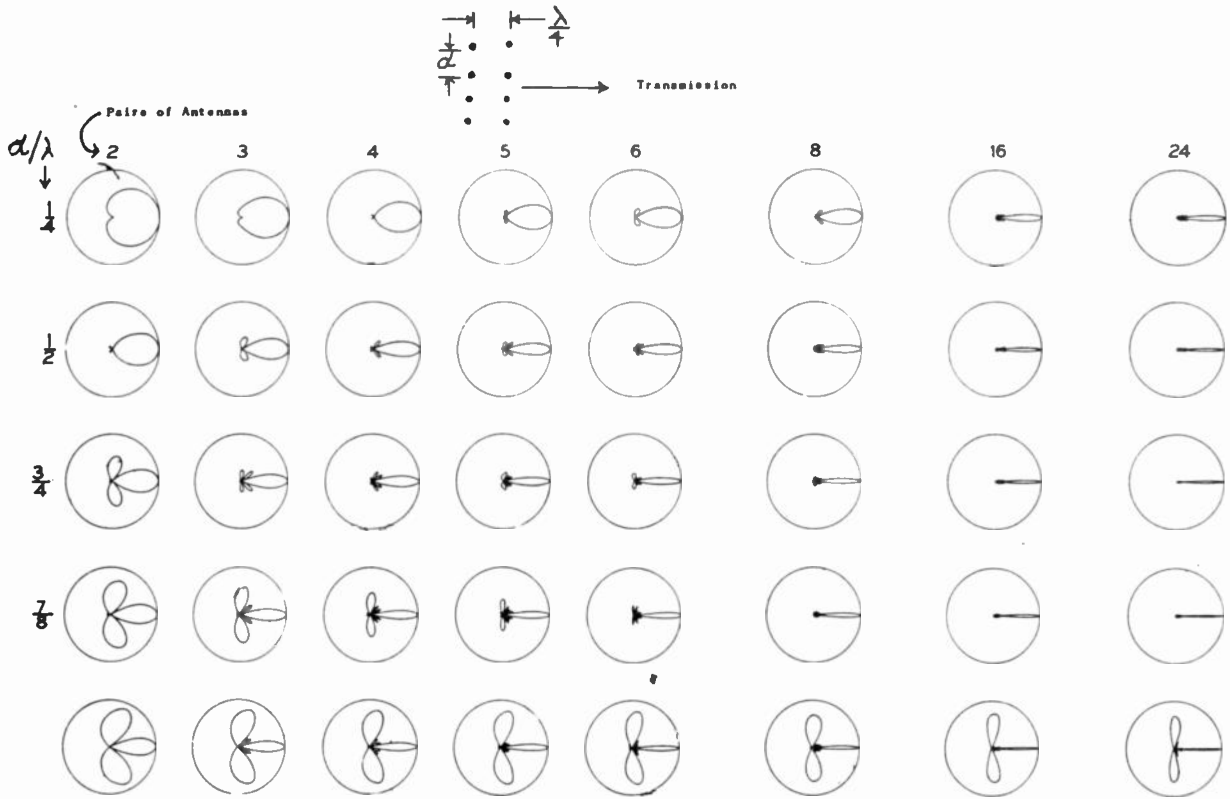


FIG. 12-15. Directional patterns of two rows of parallel antennas, separation λ/4 between rows, distance d between antennas in a row indicated at left, number of antennas in each row at top. The patterns are in a plane perpendicular to and bisecting the antennas.

$$\left.\begin{aligned} \underline{E}_1 &= \underline{Z}_{11}\underline{I}_1 + \underline{Z}_{12}\underline{I}_2 \\ 0 &= \underline{Z}_{12}\underline{I}_1 + \underline{Z}_{22}\underline{I}_2 \end{aligned}\right\} \qquad (12\text{-}98)$$

where $\underline{E}_1$ is the vector voltage applied to antenna 1, $\underline{Z}_{11}$ is the self-impedance of antenna 1, $\underline{Z}_{22}$ that of the parasitic antenna 2, and $\underline{Z}_{12}$ the mutual impedance between the antennas.. $\underline{I}_1$ is the input current of antenna 1, and $\underline{I}_2$ current at a corresponding point in antenna 2. The instantaneous current in antenna 1 causes radiation; the induced current in antenna 2 likewise causes radiation. Because of the phase differences between $\underline{I}_1$ and $\underline{I}_2$, simple spacings between antennas do not result in best effects. Figure 12-16 shows radiation patterns in a plane normal to, and bisecting, two quarter-wave antennas, and serves to illustrate a few of the many results which can be achieved with part driven, part parasitic, arrays.

12-22. Other Antenna Systems.- This chapter is intended to introduce the reader to antenna systems, and mention can be made of only a few types other than those already discussed.

Lindenblad[31] describes a "billboard" type antenna consisting of vertical half-wave antennas arranged in horizontal rows with the rows one above the other. A flat sheet-metal reflector is used behind the array, so that there is virtually no back radiation. By the use of images described previously, the theoretical patterns of the antenna system can be determined without excessive effort.

The turnstile antenna[32] consists of two linear antennas, lying in a plane, and crossing one another (without electrical contact) at their centers. If the angle between the two is 90°, the length of each half a wavelength, and the currents 90° out of phase, the directional pattern in the plane of the two is closely a circle. This can be shown thus: let H_1 be the field due to antenna 1, and H_2 that due to 2; then using the approximate formula (12-78),

$$H_1 = C\,\frac{\sin\theta}{r}\,\sin(\omega t - \beta r) \qquad (12\text{-}99)$$

$$H_2 = C\,\frac{\cos\theta}{r}\,\cos(\omega t - \beta r) \qquad (12\text{-}100)$$

$$H_1 + H_2 = C\,\frac{\sqrt{\sin^2\theta + \cos^2\theta}}{r}\,\cos(\omega t - \beta r - \theta) \qquad (12\text{-}101)$$

31. Lindenblad, N.E., Development of Transmitters for Frequencies above 300 Megacycles, Proc. I.R.E., 23, No. 9, Sept. (1935).

32. Brown, George H., The Turnstile Antenna, Electronics, 9, No. 4, April (1936)

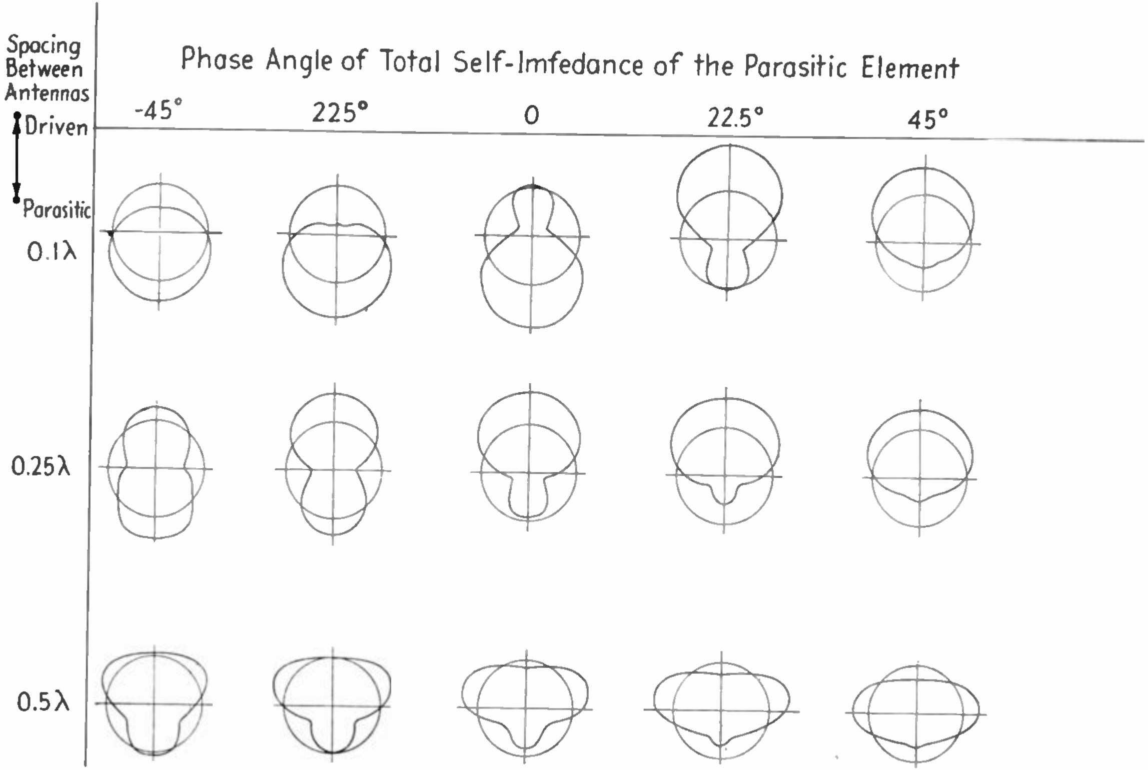


FIG. 12-16. Horizontal patterns of two vertical quarter-wave antennas, one driven and one parasitic. Distances between antennas are indicated at left, phase angle of the self-impedance of the parasitic antenna at top. Note that for the tuned case (phase angle zero), maximum radiation is in the direction of the parasitic antenna, it is then a "director"; whereas, for a phase angle of +22.5°, the converse is true and the parasitic antenna is then calked a "reflector." (From G.H. Brown, Proc. I.R.E., Jan. 1937.)

whence $$F(\theta) = \sqrt{\sin^2\theta + \cos^2\theta} = 1. \qquad (12\text{-}102)$$

In the above, $C = 0.023\sqrt{P_r}$; see eq. (12-78).

Turnstile antennas, composed of shaped rather than straight-wire radiators, have been used for television[33] broadcasting and are important in u-h-f work. The shaping of the antennas is to obtain, among other objectives, broad tuning (see Sec. 12-19).

Loop antennas are of considerable importance. The loops are usually of square shape, although some receiving loops (as in marine direction finding) are circular. Restricting discussion to square loops in which the magnitude of the current is approximately constant throughout, in the (horizontal) plane of a loop, the pattern is substantially a circle (compare the horizontal pattern of a single vertical linear antenna), and in the vertical plane the pattern is substantially a circle tangent to the vertical axis (compare the vertical pattern of the short-length vertical linear antenna). To increase horizontal radiation, loops are sometimes stacked $\lambda/2$ apart, one above the other.

Binomial arrays are loop combinations in which the magnitude of the current varies from loop to loop in proportion to the binomial coefficients. The vertical pattern is given by

$$2^{n-1}\sin\theta\,[\cos^{n-1}(a\cos\theta)] \qquad (12\text{-}103)$$

where $a = \pi d/\lambda$. Some patterns are given by Kandoian, loc. cit.

V antennas are antennas in the shape of a V. Sometimes, two parallel two-wire lines, each in a V, are used. Good directivity and gain may be obtained. In the case of the V antenna, the angle between the two arms is controllable. Carter, Hansell, and Lindenblad[34] give theory and experimental results for this type of antenna.

The rhombic antenna is so called because of its rhombic or diamond shape. If the input is at one of the four vertices, a resistor is usually inserted at the opposite vertex, and if the resistor approximates the characteristic impedance, standing waves are minimized. The amplitude of the current is then roughly uniform. Rhombic antennas are in wide use for certain purposes.[35]

The Franklin antenna is the equivalent of a number of half-wave linear antennas, all lying end to end along a straight line, and all excited in phase. This result is achieved by

---

33. Lindenblad, N.E., Television Transmitting Antenna for Empire State Building, RCA Review, 3, No. 4, April (1939).
34. Carter, P.S., Hansell, C.W., and Lindenblad, N.E., Development of Directive Transmitting Antennas by R.C.A. Communications, Inc., Proc. I.R.E., 19, No. 10, Oct. (1931).
35. Harper, A.E., "Rhombic Antenna Design," D. Van Nostrand Co., 1941.

taking a long linear antenna and, at each half-wave point, "folding" the antenna in such a way that no effective radiation is obtained from alternate half-wave sections of the line. Another method is to insert inductance, or wind the antenna wire into a coil at half-wave points, in such a way that the same result is obtained. The current distribution thus appears as a standing wave each half wavelength of which is in phase with all other half wavelengths. (In a linear antenna the successive half wavelengths of the current standing wave are in opposite phase.) If a number of equal Franklin antennas are placed parallel to and at uniform distances from one another, a billboard type array (not including the sheet reflector) results. The radiation resistance of a Franklin antenna is approximately 100 m ohms where m is the number of half-wave loops in the standing wave pattern ($m > 1$). The radiation pattern of a vertical linear half-wave antenna with bottom on a perfect conductor (earth) is the same as the pattern for the Franklin antenna for m = z. This pattern has already been given. The formula for m loops is[36]

$$F(\theta) = \frac{\cos(\frac{\pi}{2} \cos \theta) \sin (\frac{m\pi}{2} \cos \theta)}{\sin \theta \sin(\frac{\pi}{2} \cos \theta)} \qquad (12\text{-}104)$$

12-23. Reflection.- Up to this point the passage of electromagnetic waves in a single medium, such as an ideal non-conducting atmosphere about the earth, has been considered. Mention was made of the presence of a second medium in the case in which the effect of the ground must be taken into account in determining radiation in the space above it, and by making the assumption that the ground was the equivalent of a perfect conductor, it was possible to circumvent the two-medium problem through the use of a fictitious image antenna in one medium.

When an electromagnetic wave passing through one medium strikes another medium, which may be of large size such as the earth, or may be of small size, the resultant effect in the first medium and the effect produced in the second medium depend very much on the dimensions of this second medium, as well as on its shape, electric and magnetic properties, etc.

We will consider in this section the simplest case which can be set up and one which is at the same time important because it approximates rather closely numerous practical problems, and indicates the type of result to be anticipated in others. This is the case of a plane wave traveling through a non-conducting medium and striking a second non-conducting[37] medium,

36. Stratton, loc. cit.

37. Note that the assumption of a non-conducting second medium makes the problem not analogous to a wave passing through the atmosphere and striking the ground, but roughly analogous to a wave passing through an ideal non-conducting atmosphere and striking an ideal non-conducting medium such as a body of pure water (not sea water).

the surface of separation between the two media being an indefinite plane[38] which can be taken as the YZ plane (Fig. 12-17). Each medium is assumed of indefinite extent[39] in the Y and Z directions, and extending from 0 to ∞ in the + X direction (medium 1), or from 0 to $-\infty$ in the - X direction (medium 2). Furthermore, it is desirable to point out specifically that the dielectric constant and the permeability of each medium are assumed to be constant[40] (ε_1 and μ_1 for medium 1, ε_2 and μ_2 for medium 2);

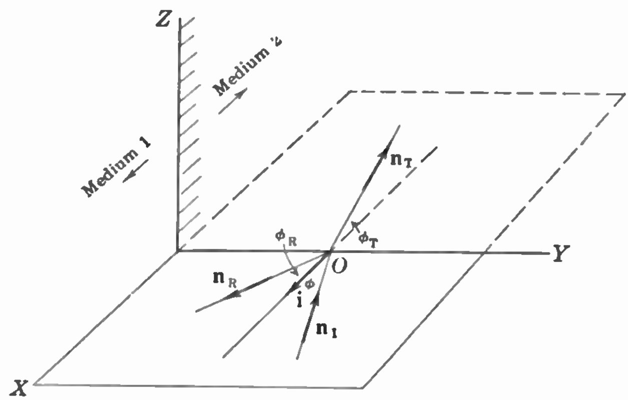


FIG. 12-17. Reflection and refraction of a plane electromagnetic wave; $\mathbf{n}_I$ is the normal to the wave-front of the incident wave, $\mathbf{n}_R$ to the reflected wave, $\mathbf{n}_T$ to the transmitted wave.
X is perpendicular to plane separating the two media; Y and Z lie in this plane. Y is chosen so that XY contains $\mathbf{n}_I$, that is, XY is the plane (plane of incidence) containing the normal (i) to the surface of separation, and the normal $\mathbf{n}_I$ to the wave-front of the plane incident wave. (Having fixed X $\perp$ to plane of separation, it is always possible to rotate Y in that plane so that XY contains $\mathbf{n}_I$.)

38. The assumption of a smooth plane surface between the two media means that a sharp transition from one medium to the other is assumed. No "smooth" flat surface is in reality of infinite smoothness, but the assumption introduces no error of consequence for our purposes.
39. The assumption of infinite extent of the surface between the two media is made for two purposes: (1) it assures that each dimension of the surface of separation is large; and, (2) that no edge effects need be considered. If either of these conditions is not satisfied, the problem becomes much more complex, and the following analysis ceases to hold. The problem becomes a "diffraction" problem rather than one of simple reflection and transmission.
40. In the event one or more of these quantities is not a constant, the analysis following does not apply. When masses of air are unevenly heated, the equivalent of this effect is obtained, and the wave which passes into the heated air (second medium) from the uniform air (first medium) may be changed in direction in such a manner that, in some cases, the wave may follow a path which causes it to emerge from medium 2 back into medium 1.

in addition, the conductivities σ_1, σ_2 of the two media are assumed to be zero, so that there is no conduction current in either medium ($J_1 = J_2 = 0$).

The problem can now be set up thus: In medium 1 it will be assumed on the basis of experience that two plane waves called the incident and reflected waves exist simultaneously at each point, and that the normals $\underline{n}_I$ and $\underline{n}_R$ to the wave-fronts are not parallel. Each wave separately must be such that $\underline{E}$ and $\underline{H}$ satisfy conditions of the type

$$\nabla^2\underline{E} = \frac{1}{c_1^2}\frac{\partial^2\underline{E}}{\partial t^2} \text{ and } \nabla \cdot \underline{E} = 0 \qquad (12\text{-}105)$$

where $c_1^2 = 1/\mu_1\varepsilon_1$. The total fields $\underline{E}_1$ and $\underline{H}_1$ at any point will be the sum of the vector fields $\underline{E}_{1I}$, $\underline{E}_{1R}$, etc. due to the two waves separately ($\underline{E}_1 = \underline{E}_{1I} + \underline{E}_{1R}$, etc.).

In medium 2, a single plane wave, called the transmitted wave, will be assumed to exist. The normal to its wave-front $\underline{n}_T$ need not necessarily be parallel to $\underline{n}_I$ or $\underline{n}_R$.

The YZ-axes will lie in the plane separating the two media, and X will be normal to this plane in such a way that $\underline{n}_I$ lies in the XY plane. It is always possible, no matter at what angle $\underline{n}_I$ meets the separating plane YZ, to move and turn the YZ-axes in that plane until the XY plane contains $\underline{n}_I$. Figure 12-17 shows the arrangement of axes. The transmitted wave must be such that the fields $\underline{E}_2$ and $\underline{H}_2$ in medium 2 satisfy at every point in medium 2, and at every time, equations of the form ($c_2^2 = 1/\mu_2\varepsilon_2$)

$$\nabla^2\underline{E}_2 = \frac{1}{c_2^2}\frac{\partial^2\underline{E}}{\partial t^2} \text{ and } \nabla \cdot \underline{E}_2 = 0 \qquad (12\text{-}106)$$

Now if the X-axis were parallel to $\underline{n}_I$, and thus perpendicular to the wave-front of the incident wave, $\underline{E}_I$ would have components of the form A cos $(\beta x - \omega t)$. Since X is not perpendicular to the wave-front, but $\underline{n}_I$ is, $\underline{E}_I$ will have components of the form

$$E_{Ix} \cos(\beta n_I - \omega t)$$

$$E_{Iy} \cos(\beta n_I - \omega t + \delta_y)$$

$$E_{Iz} \cos(\beta n_I - \omega t + \delta_z)$$

where δ_y and δ_z are constant, and n_I is distance from 0 to the wave-front. If x, y are the coordinates of any point on $\underline{n}_I$,

$$n_I = x\, C_{xI} + y\, C_{yI} \qquad (12\text{-}107)$$

where C_{xI} is the cosine of the angle between X and $\underline{n}_I$, etc. ($C_{xI} = \cos\varphi$; $C_{yI} = \sin\varphi$; φ = angle between X and $\underline{n}_I$ as shown in Fig. 12-17.) Likewise,

$$n_R = x\, C_{xR} + y\, C_{yR} + z\, C_{zR} \tag{12-108}$$

$$n_T = x\, C_{xT} + y\, C_{yT} + z\, C_{zT} \tag{12-109}$$

recalling that, unlike $\underline{n}_I$, $\underline{n}_R$ and $\underline{n}_T$ do not necessarily lie in the XY plane.

At the boundary between the two media, that is, at any and every point for which $x = 0$, the following relations must hold:

1. The component of the total electric field $\underline{E}_1$ in medium 1, tangent to the surface (YZ plane) separating the two media, must equal the component of $\underline{E}_2$ tangent to YZ. This is usually written

$$E_{1tan} = E_{2tan} \tag{12-110}$$

2. The component of the total electric flux density $\underline{D}_1 (= \varepsilon_1 \underline{E}_1)$ in medium 1, normal to the surface (YZ plane) separating the two media, must equal the component of $\underline{D}_2$ normal to YZ. This is usually written $\varepsilon_1 E_{1n} = \varepsilon_2 E_{2n}$, and in the present case becomes

$$\varepsilon_1 E_{1x} = \varepsilon_2 E_{2x} \tag{12-111}$$

where E_{1x} is the X component of $\underline{E}_1$, etc.

3. $H_{1tan} = H_{2tan}$ (12-112)

4. $\mu_1 H_{1n} = \mu_2 H_{2n}$ or, here, $\mu_1 H_{1x} = \mu_2 H_{2x}$ (12-113)

These four laws, commonly referred to as boundary conditions because they describe relations which must obtain at the boundary between two media, are developed in elementary texts on electricity and magnetism, and hold equally well at high as well as at low frequencies. They are of fundamental importance here, because they determine the directions of $\underline{n}_R$ and $\underline{n}_T$ with reference to $\underline{n}_I$, the magnitudes of the R and T waves in terms of that of the I wave, and other results for the problem.

Boundary condition 2 in the present case states that $\varepsilon_1 E_{1x} = \varepsilon_2 E_{2x}$ when $x = 0$, that is

$$\varepsilon_1 E_{Ix} \cos(\omega t - \beta y \sin \varphi) + \varepsilon_1 E_{Rx} \cos[\omega t - \beta(yC_{yR} + zC_{zR}) + \delta_R]$$
$$= \varepsilon_2 E_{Tx} \cos[\omega t - \beta_T(yC_{yT} + zC_{zT}) + \delta_T] \tag{12-114}$$

and this equality must hold for any and every value of t, y, and z. The only way in which this can be realized is to have the coefficient of t, of y, and of z in each argument equal the corresponding coefficients in the other two arguments of the cosines, that is,

$$\beta \sin \varphi = \beta\, C_{yR} = \beta_T\, C_{yT} \tag{12-115}$$

and

$$0 = \beta\, C_{zR} = \beta_T\, C_{zT} \tag{12-116}$$

Thus, $C_{zR} = C_{zT} = 0$, which means that $\underline{n}_R$ and $\underline{n}_T$ lie in a plane (the XY plane) at 90° to Z, and

$$\varphi_R = \varphi \tag{12-117}$$

since $\sin\varphi = C_{yR}$, and

$$\frac{C_{yR}}{C_{yT}} = \frac{\beta_T}{\beta} = \frac{\sin\varphi}{\sin\varphi_T} = \frac{c_2}{c_1} = \frac{\lambda}{\lambda_T} \tag{12-118}$$

Equation (12-117) states that the angle of reflection, φ_R, is equal to the angle of incidence, φ. Equation (12-118) gives φ_T, the angle which the perpendicular to the wave-front of the transmitted wave makes with the X-axis, in terms of φ and the velocities of phase propagation c_1 and c_2 in medium 1 and medium 2, respectively. The wavelength λ_T of the transmitted wave is obtained from $\beta_T = 2\pi/\lambda_T = \omega/c_2$.

The boundary conditions apply to components normal to the surface of separation (YZ plane here) and tangential to this surface. It is convenient to consider the field vectors E and H broken down into two components, one normal to XY, which is the usual Z component, and one tangent to XY, which is the resultant E' of the X and Y components. Figure 12-18 illustrates this breakdown and shows the relation between the E' and kE_z mentioned here and the usual XYZ components.

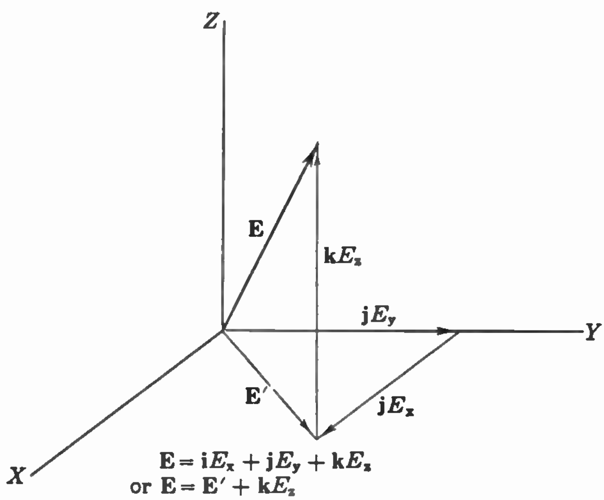


FIG. 12-18. E expressed as the sum of three rectangular components, and as the sum of a component kE_z normal to XY and a component E parallel to XY.

A. Components of E vectors normal to XY. In this case, in which the E vectors are usually said to be polarized normal to the plane of incidence XY since, taken alone, the components normal to XY form such waves,

$$\left.\begin{array}{ll} & E_{Iz} + E_{Rz} = E_{Tz} \\ \text{and} & H_{Iz}\cos\varphi - H_{Rz}\cos\varphi = H_{Tz}\cos\varphi_T \\ \text{or} & (E_{Iz} - E_{Rz})\cos\varphi = \frac{\beta\mu_2}{\beta_2\mu_1} E_{Tz}\cos\varphi_T \end{array}\right\} \text{at } x = 0 \qquad (12\text{-}119)$$

whence,

$$\frac{E_{Rz}}{E_{Tz}} = \frac{1 - q}{2} \quad \text{and} \quad \frac{E_{Iz}}{E_{Tz}} = \frac{1 + q}{2} \qquad (12\text{-}120)$$

where $q = \beta\mu_2 \cos\varphi_T/\beta_2\mu_1 \cos\varphi$. Equations (12-120) give the magnitudes of the Z components of the three waves (I, R, T) in terms of any one. Since

$$\beta \cos\varphi_T = \sqrt{\beta^2 - \beta_2^2 \sin^2\varphi}$$

the factor q appearing in (12-120) can be expressed entirely in terms of φ.

B. Components of $\underline{E}$ vectors parallel to XY. In this case the waves are usually said to be polarized in the plane of incidence XY. When the $\underline{E}$ vector lies in the plane XY, or when only the component of $\underline{E}$ in the plane of incidence XY is considered, there is an accompanying $\underline{H}$ field perpendicular to $\underline{E}$ (see earlier discussion of plane waves) and hence perpendicular to XY, that is, in the Z direction. Since $\underline{E}$ can be found when $\underline{H}$ is known and vice versa, we can deal with the $\underline{H}$ vector in this case. This is a very convenient procedure to follow, because $\underline{H}$ now occupies the same position as $\underline{E}$ in part A, and very nearly similar boundary conditions hold for $\underline{H}$ in part B as for $\underline{E}$ in A. Indeed, since $\underline{H}$ is in the Z direction, hence a tangent to YZ,

$$\left.\begin{array}{ll} & H_{Iz} + H_{Rz} = H_{Tz} \\ \text{and} & H_{Iz}\cos\varphi - H_{Rz}\cos\varphi = \frac{\mu_1\beta_2}{\mu_2\beta_1} H_T \cos\varphi_T \end{array}\right\} \text{at } x = 0 \qquad (12\text{-}121)$$

whence,

$$\frac{H_{Iz}}{H_{Tz}} = \frac{Hq'}{2} \quad \text{and} \quad \frac{H_{Rz}}{H_{Tz}} = \frac{1 - q'}{2} \qquad (12\text{-}122)$$

where $q' = \mu_1^2 q/\mu_2^2$. Thus, H_{Tz} can be found in terms of H_{Iz}, then H_{Rz}, and the magnetic fields of the I, R, and T waves determined. The $\underline{E}$ fields follow from

$$\underline{E} = -\frac{\omega\mu}{\beta}\, \underline{n} \times \underline{H} \qquad (12\text{-}123)$$

where any one of the subscripts I, R, T is to be placed on $\underline{E}$, $\underline{n}$, and $\underline{H}$.

The objective of the above discussion is to introduce the reader to a small part of the theory reflection and refraction of electromagnetic waves. If the second medium were a conductor, the transmitted wave would be attenuated rapidly, the energy going into heat. But, as the conductivity becomes better, more energy is reflected, until in the limit all incident energy is reflected by a perfect conductor.

If the reflecting surface were not flat, or not large, compared with the wavelength, the results obtained above would not hold. Problems dealing with objects of the same order of magnetude as the wavelength of incident radiation, or with edge effects, passage of radiation through apertures, etc., are diffraction problems. It is not possible to go into these here, but diffraction effects are important in connection with thé subject of the next section.

Flat sheets of metal form good reflectors provided they are large enough. Mention was made of such reflectors in connection with several antenna systems. They are used chiefly to cut down back radiation and enhance that in the forward direction.

12-24. The Paraboloid Reflector.- If an antenna is placed at the focus of a sheet or block of metal whose inner surface is a parabola of revolution, the radiation is to a large extent directed along the axis of the system. Unfortunately, however, the antenna radiation is not symmetrical about the focus, the wavelength may be not negligible in comparison with paraboloid dimensions, and in general a diffraction problem of considerable theoretical difficulty is presented. The radiation pattern is not a very thin lobe along the axis, but is a lobe of appreciable cross section, accompanied by minor lobes. The reader who carries out the experiment described in Sec. 15-16 will learn more than can be described in the space available here.

In Ch. 4, horns are mentioned, and the experiment of Sec. 15-16 includes a simple investigation of horn radiation. The electromagnetic energy is conducted to a horn by a wave guide, and the horn then serves to give a direction to the radiation in space. The horn, however, presents another diffraction problem which has for practical purposes been solved. It is mentioned here because the horn and the paraboloid represent two sides of a sharply defined transition point. In the case of the paraboloid, the antenna is essential and must be near the focus for effective directivity. In the horn, the antenna is not essential, and if used its position is not critical. If a wave guide is used to lead energy to the horn, the antenna is effectively removed. On the other hand, the extension of the paraboloid surface makes the paraboloid approach a wide-aperture horn of large flare angle. It is thus seen that the horn and the paraboloid represent an approach to a meeting of two different radiation methods.

Chapter 13

PROPAGATION

In Ch. 12 the radiation produced by various antenna systems in a homogeneous isotropic perfectly insulating medium free of electric charge was discussed. In this chapter we disregard the source of the radiation and review briefly the properties and the effects of the actual medium through which the energy passes.

The earth's magnetic field is a commonplace, and the earth's electric field, although less well known to the average person, is an important electrical characteristic of the space about the earth. In the atmosphere, which may be taken to be the region containing the air about the earth, there are sundry electrical phenomena such as lightning with which everyone is familiar. The troposphere is that part of the atmosphere below approximately 11 km elevation in which the temperature varies appreciably with altitude. In the troposphere there are many causes of non-uniformity of the medium, such as sharp temperature changes in short distances, and this lack of homogeneity sometimes appreciably influences the passage of electromagnetic waves. The stratosphere is that portion of the atmosphere above the troposphere and below the ionosphere in which the temperature is almost constant. Above the stratosphere lies the ionosphere, "that portion of the earth's atmosphere above the lowest level at which the ionization is large compared with that at the ground, so that it affects the transmission of radio waves." The bottom of the ionosphere is not less than approximately 50 km.

The reader will now have an appreciation of the fact that the atmosphere considered as a medium in which electromagnetic waves are propagated may be decidedly non-homogeneous, and may contain electric charges as is evident in the case of the ionosphere. If we look at the region immediately below the atmosphere, that is, the earth or ground, we find a medium which often appears as a good conductor for waves of not very high frequencies, and gradually comes to appear as a dielectric at ultra-high frequencies.

13-1. Ground Wave, Sky Wave, etc.- Turning now to an inspection of the radiated wave, we may distinguish three separate major components in the most general case, a ground wave, a sky wave (also called an ionospheric wave), and a tropospheric wave. The ground wave is "the portion of a radio wave that is propagated through space and is, ordinarily, affected by the presence of the ground." It is the total received wave less the ionospheric and tropospheric waves.

The sky wave (ionospheric wave) is the part of a radio wave that has been directed toward, or reflected from, the ionosphere.

The tropospheric wave is any part of a radio wave that has been reflected from a place of abrupt change of dielectric constant in the troposphere (such as the boundary between hot and cold layers).

The ground wave[1] is usually further subdivided into a direct wave,[2] a wave reflected[3] from the ground (of importance when the receiving antenna is well above the ground). and a surface or guided wave. The ground wave is usually refracted in passing through the lower atmosphere, and this combined with the guiding effect which exists (the earth may act as a wave guide somewhat as one wire of a transmission line does) tends to cause the ground wave to follow the curvature of the earth when the frequency is not too great. But the ground wave often suffers severe attenuation, so that it cannot often account for long-distance transmission except at relatively low frequencies.

13-2. Factors Affecting Propagation.- With the few paragraphs above--which do little more than indicate the names of a few of the characters met in the study of radio wave propagation--to introduce the topic, some of the more important factors which influence propagation and reception may be enumerated:

1. Frequency
2. Effect of ground
3. Curvature of the earth
4. Homogeneity of the troposphere

---

1. "A ground wave is the portion of a radio wave that is propagated through space and is, ordinarily, affected by the presence of the ground. The ground wave includes all components of the received waves except ionospheric waves and tropospheric waves. The ground wave is somewhat refracted by the gradual change of the dielectric constant of the lower atmosphere with altitude."

2. "A direct wave is the component of the ground wave that is transmitted directly from transmitter to receiver. It is the entire ground wave where the transmission path is relatively distant from the ground, as in transmission from one airplane to another at a distance small compared to their distance from the ground. Where the transmission path is along or near the ground, the direct wave is that component of the ground wave which would exist if the ground were not present. Where the path of transmission is obstructed, as by the bulge of the earth, the direct wave is absent."

3. "A ground-reflected wave is the component of the ground wave that is reflected from the ground according to the laws of geometric optics."

5. Time of day, time of year, time in the solar cycle
6. Characteristics of the ionosphere
7. Earth's magnetic field
8. Local disturbances such as thunderstorms, which result in the introduction of noise, and distant disturbances such as magnetic storms of the sun, and ionospheric storms

This list does not include any mention of radiating or receiving systems, and is the only mention to be made of disturbed conditions (item 8) as contrasted with the non-disturbed state. Furthermore, we shall make no further reference to the troposphere, since the modifications in wave propagation caused by it are in part erratic and in part not feasible or important enough to handle here. But the height of the transmitting and receiving antennas above ground may be a substantial consideration.

Frequency is a vital differentiator between the importance of the numerous effects. Following Mimno,[4] the frequency range of interest may be divided into six sections for convenience, although it should be understood that no sharp and fast line of demarcation holds between any two adjacent divisions:

Range	Frequency (f)	Wavelength (λ) in meters
I	50 - 550 kc	6000 - 545
II	550 - 1500 kc	545 - 200
III	1.5 - 5 Mc	200 - 60
IV	5 - 30 Mc	60 - 10
V	30 - 300 Mc	10 - 1
VI	300 - 3000 Mc	1 - 0.1

We shall return to these frequency classifications later. To discuss in detail transmission in each range is not possible here; to discuss it even briefly requires some description of the ionosphere, also known as the Heaviside or the Kennelly-Heaviside layer.

13-3. The Ionosphere.- In Fig. 13-1 is shown a graph of possible variation of relative electron density versus altitude. Because of the sun's radiation, cosmic rays, and other causes, it is well determined experimentally that the region now known as the ionosphere contains ions, those of interest here being chiefly free electrons, in a number per unit volume which depends on many factors including height and time of day, time of year, and time in the sun-spot cycle. There is a definite stratification of the ionosphere, as indicated by peaks E,

4. Mimno, H.R., The Physics of the Ionosphere, Rev. of Mod. Physics, 9, No. 1, Jan. (1937). Mimno's frequency sections are slightly different from those given here.

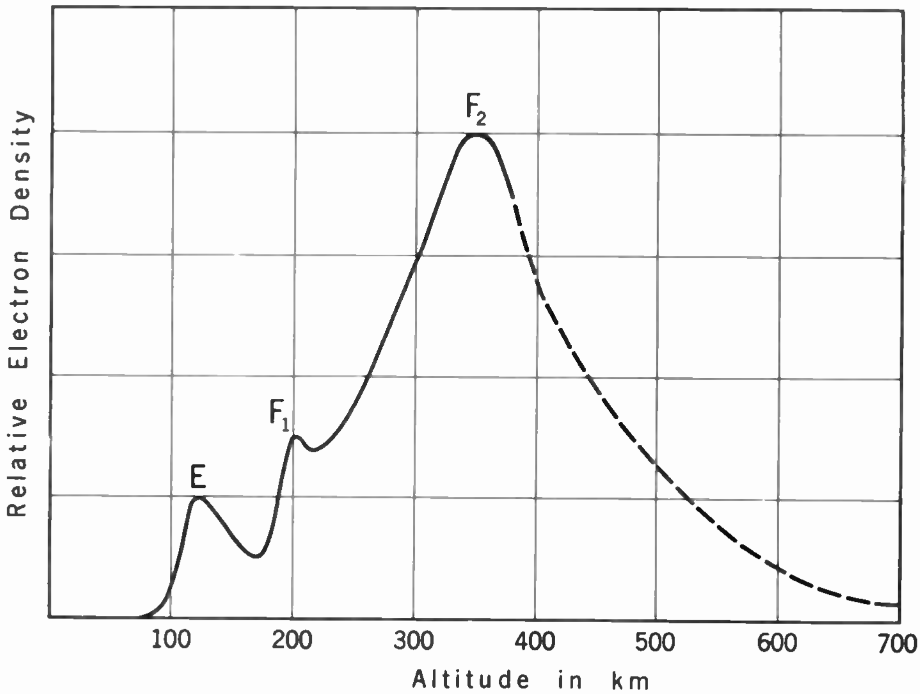


FIG. 13-1. Possible Variation of electron density with altitude above 80 km.

F, and F_2 in Fig. 13-1, so that layers--"a layer is a regularly stratified distribution of ionization which is formed in a region of the ionosphere and which is capable of reflecting radio waves back to earth"--are recognized. The D layer is not as well defined and for most purposes not as important as other layers. It is not shown on Fig. 13-1, but would, if present, be represented by a small peak left of E.

The E layer is a layer somewhere in the E region, which is the region between approximately 90 and 140 km above the surface of the earth. The F layer is one which exists in the F region (140 to 400 km above the surface of the earth) "in the night hemisphere and in the weakly illuminated portion of the day hemisphere." Over most of the day hemisphere two layers, F_1 and F_2, exist rather than one F layer, and these two have been indicated on Fig. 13-1. The F_1 layer is the lower and the F_2 layer the higher of the two day layers in the F region. The virtual heights of the various layers can be measured and Fig. 13-2 has been inserted to give some idea of the variation of position of the various layers during a year.

Although it is not possible here to go into the mechanism of the creation of the ionization in the ionosphere, it is possible to touch on certain electrical effects, assuming that ionization to exist.

In Ch. 12 a section was devoted to the discussion of displacement current, and it was pointed out that the total current density $\underline{J}_T$ (amperes per square meter) is

$$\underline{J}_T = \underline{J} + \partial \underline{D}/\partial t \tag{13-1}$$

where $\underline{J}$ is the conduction current density, $\underline{D}$ is the electric flux density, and $\partial \underline{D}/\partial t$ is the displacement current density. Since $\underline{D} = \varepsilon \underline{E}$,

$$\underline{J}_T = \underline{J} + \varepsilon \partial \underline{E}/\partial t \tag{13-2}$$

Now assume an electric field $\underline{E}$ having the same direction at all points (such as the field between the plates of a parallel plate capacitor except near the plate edges), and let $E = E_0 \sin \omega t$. Then

$$\varepsilon \partial E/\partial t = \varepsilon \omega E_0 \cos \omega t \tag{13-3}$$

If there are free electrons present in an otherwise empty medium or its equivalent, $\rho = -ne$ where n is the number of electrons per unit volume and e is the magnitude of the negative charge on one electron. The conduction current density J is ρv (charge crossing unit area in unit time; corresponds to dq/dt in a wire circuit) where v is the velocity of motion of the electrons. But if the electric field $E = E_0 \sin \omega t$ is the only force effectively acting on the electrons, then by Newton's law

$$m \frac{dv}{dt} = -e E_0 \sin \omega t \tag{13-4}$$

or

$$v = \frac{eE_0}{m\omega} \cos \omega t \tag{13-5}$$

whence

$$\rho v = -\frac{ne^2}{m\omega} E_0 \cos \omega t \tag{13-6}$$

and

$$J_T = -\frac{ne^2}{m\omega} E_0 \cos \omega t + \varepsilon \omega E_0 \cos \omega t$$

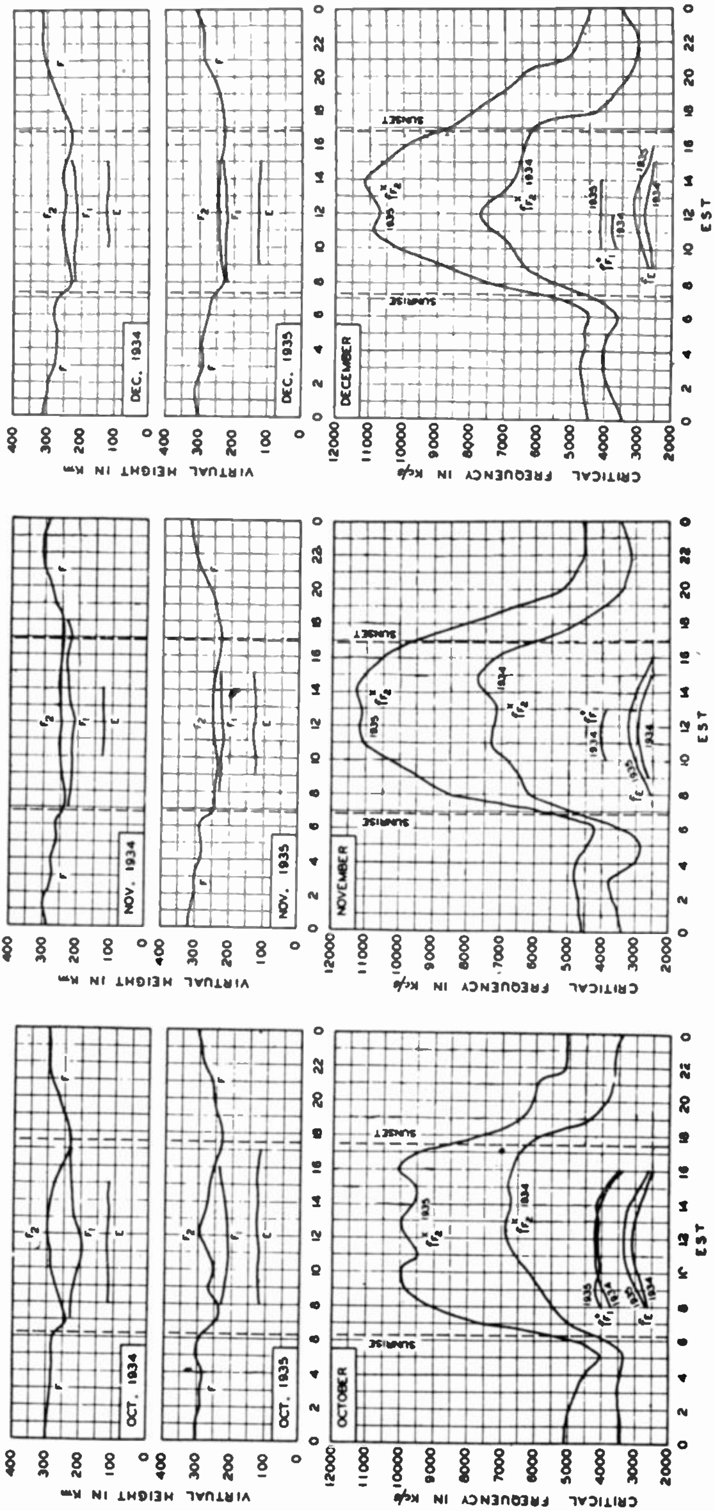
or

$$J_T = \left(\varepsilon - \frac{ne^2}{m\omega^2}\right) \frac{\partial E}{\partial t} \tag{13-7}$$

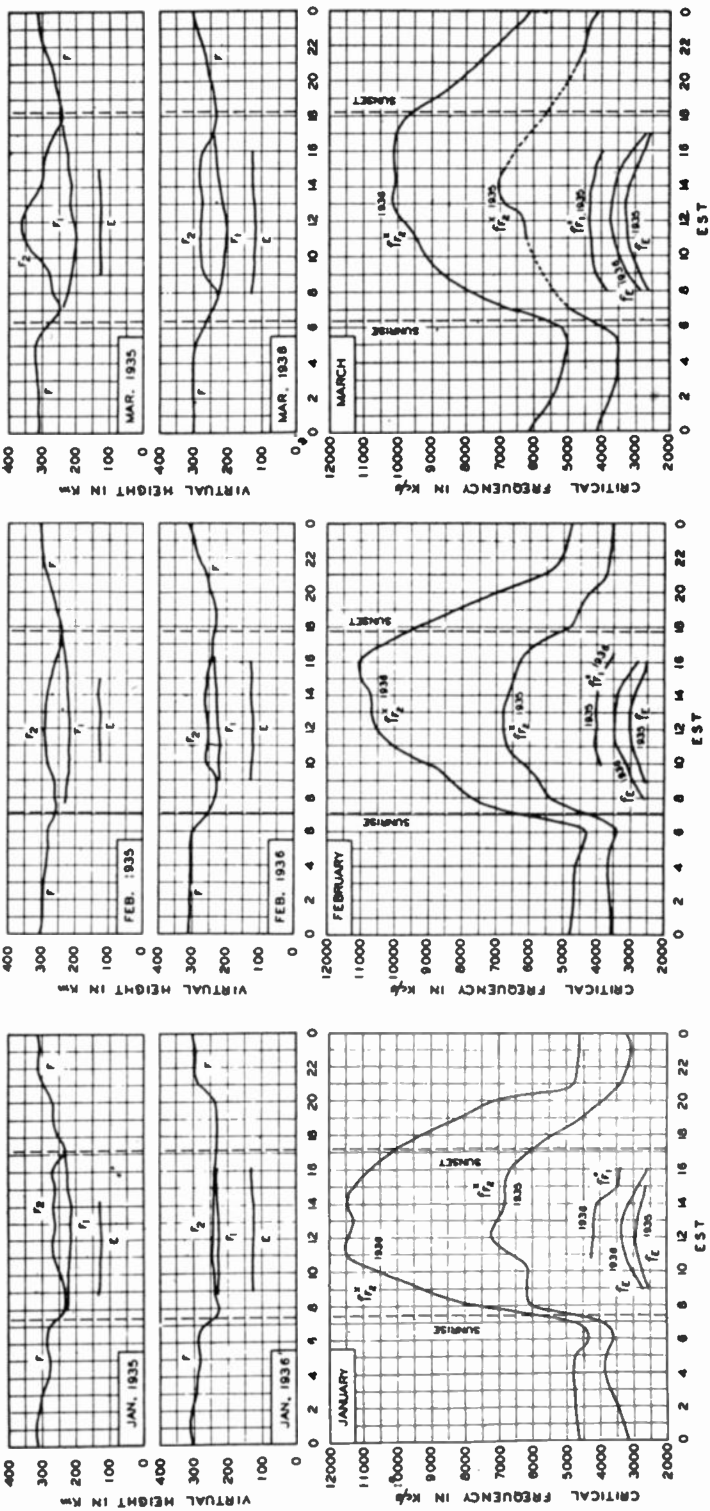
Thus the ionized medium acts like a dielectric medium of dielectric constant

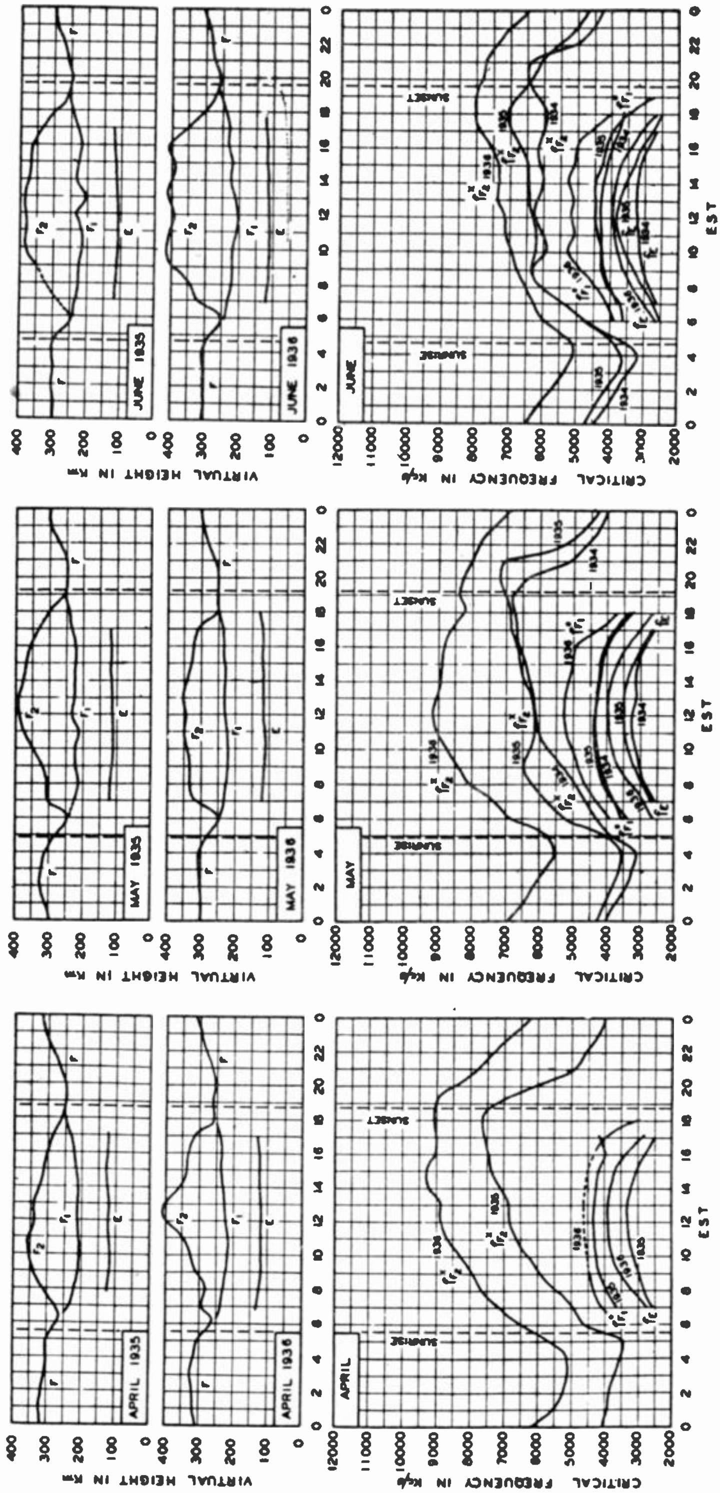
$$\varepsilon - \frac{ne^2}{m\omega^2} = \varepsilon\left(1 - \frac{ne^2}{m\omega^2 \varepsilon}\right) = \varepsilon\left(1 - \frac{f_c^2}{f^2}\right) \tag{13-8}$$

where $f = \omega/2\pi$ is the frequency of the applied electric field and



EST - Eastern Standard Time


MAR. 1935
MAR. 1936
MARCH
FEB. 1935
FEB. 1936
FEBRUARY
JAN. 1935
JAN. 1936
JANUARY
SUNRISE
SUNSET
VIRTUAL HEIGHT IN Km
CRITICAL FREQUENCY IN Kc/s
EST


JUNE 1935
JUNE 1936
JUNE
MAY 1935
MAY 1936
MAY
APRIL 1935
APRIL 1936
APRIL
VIRTUAL HEIGHT IN Km
CRITICAL FREQUENCY IN Kc/s
EST
SUNSET
SUNRISE

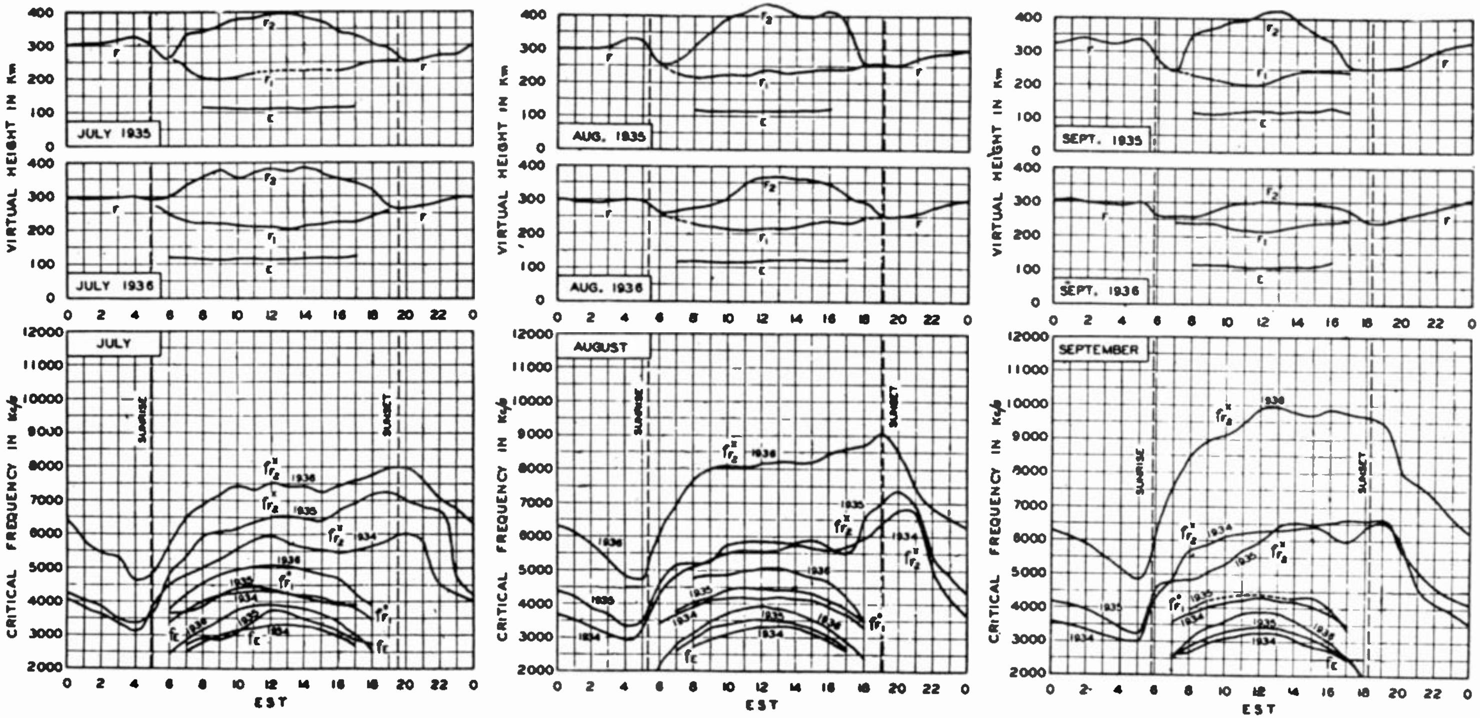


FIG. 13-2. Virtual heights and critical frequencies (hourly averages) for two years. E, F, F_1, F_2, refer to layers; f to frequency, superscript X to the extraordinary wave and O to the ordinary wave. (From Gilliland, Kirby, Smith and Reymer, "Characteristics of the Ionosphere and Their Application to Radio Transmission," Proc. I.R.E., 25, No. 7, July 1937.

$$f_c^2 = \frac{ne^2}{4\pi^2 m\varepsilon} \tag{13-9}$$

is the critical frequency, so called because the equivalent dielectric constant becomes zero when $f = f_c$. Equation (13-9) is one of the most important formulas now used in the study of the ionosphere.

Again calling on the material of Ch. 12, it will be remembered that the phase velocity in a dielectric medium is

$$c = \sqrt{1/\mu\varepsilon} \quad \text{meters per sec} \tag{13-10}$$

where μ and ε are respectively the permeability and dielectric constant of the medium.

The refractive index η of a medium is the ratio of the phase velocity of a wave in empty space to the phase velocity of the same wave in that medium. Taking the permeability of the equivalent dielectric medium mentioned above and of free space equal, and ε as the dielectric constant of empty space,

$$\eta = \sqrt{1 - \frac{ne^2}{m\omega^2\varepsilon}} = \sqrt{1 - \frac{f_c^2}{f^2}} \tag{13-11}$$

for the equivalent dielectric medium which an ionosphere layer represents. When a radio wave is directed vertically upwards [so that all vectors in (13-2) are in the same direction], the above analysis holds to a good approximation, and when $f = f_c$, the phase velocity in the upward direction becomes zero. The wave is then reflected back. From f_c, the electron density n can be found; from the time interval between release of the signal and its return, the virtual height from which it has been reflected can be found.[5] Both virtual height and critical

5. "The virtual height of a layer of the ionosphere is the height at which reflection from a sharp boundary would result in the same time of travel as that actually observed, for a wave transmitted from the ground to the ionosphere and reflected back. Consequently, it is the product of the velocity of electromagnetic waves in free space by half the time taken by a wave group to travel vertically once to the ionosphere and back to its starting point. Virtual height depends upon the wave component and frequency for which the measurement is made; the value usually stated is for the ordinary wave and for the lowest frequency used in the measurement. (This is sometimes called equivalent height.)

The penetration frequency of a layer of the ionosphere, for a wave component, is the frequency at which the virtual height for that component becomes a maximum because of penetration of the wave through the layer, for waves reflected vertically from the layer. Except for the occurrence of sporadic and scattered reflections, it is the highest frequency of waves reflected from the layer at vertical incidence. The square of the penetration frequency of the ordinary wave is proportional to the maximum equivalent electron density of the layer. (This has been called the critical frequency.)"

frequency for various hours of the day for each month of a year are shown in Fig. 13-2. The ordinary[6] and extraordinary[7] waves noted in the critical frequency curves of the figure are two waves into which a radio wave is split in the ionosphere as a result of the action of the earth's magnetic field on the electrons at the same time the electric field of the wave acts on them. The theory is given by Nichols and Schelling.[8]

The determination of virtual heights involves a technique using many of the principles discussed in this text. A pulse is generated electrically and radiated vertically upwards. At the instant it leaves the transmitter a signal is recorded on a nearby receiver in some manner. The transmitter is then

---

6. "The ordinary wave is one of the two components into which a radio wave is split in the ionosphere by the earth's magnetic field. In the lower parts of the ionosphere the ordinary wave is left-handed elliptically polarized where the earth's magnetic field has a positive component in the direction of propagation. In the special case of vertical propagation at the magnetic equator the ordinary wave is unaffected by the earth's magnetic field. For other conditions of propagation the magnetic field affects both wave components, but the name "ordinary wave" is retained as a designation of the wave component which is defined by continuity from this special case. The term is used in a different sense in optics. (This wave is designated by the letter symbol O and is sometimes called the O-wave.)"
7. "The extraordinary wave is the other of the two components into which a radio wave is split in the ionosphere by the earth's magnetic field. In the lower parts of the ionosphere the extraordinary wave is right-handed elliptically polarized where the earth's magnetic field has a positive component in the direction of propagation. (This wave is designated by the letter symbol X and is sometimes called the X-wave.)"
8. Nichols, H.W., and Schelling, J.C., Propagation of Electric Waves over the Earth, Bell Sys. Tech. Jour., 4, No. 2, April (1925). Although we cannot go into the use of the quantity here, it may be well to quote the following definition, associated with the effect of the earth's field on ions in the ionosphere: "The gyro frequency is the natural frequency of rotation of ions around the lines of magnetic field of the earth. Letting H be the total intensity of the earth's magnetic field in ampere-turns per meter, e the charge in coulombs and m the mass in kilograms of an ion, the gyro frequency in cycles per second is $f_h = 2(10)^{-7} H \frac{e}{m}$. Since H varies considerably in different geographical locations and decreases with height above the surface of the earth, f_h does also. For electrons, e/m is equal to 1.769×10^{11}, so that f_h in kilocycles per second is equal to 35.38H. Thus, f_h is of the order of 700 to 1600 kilocycles, while for other ions it is in the audio-frequency range. (Half this frequency is sometimes called the Larmor precession frequency.)"

cutoff, and the receiver, which must have recovered from the signal due to the initiation of the pulse, now receives and records the reflected pulse. The entire operation takes a time of the order of magnitude of a few milli-seconds. The time between the initiation of the pulse and the reception of the reflected signal is determined from the distance between pips on the recording device, or otherwise, and half this time multiplied by the velocity of electromagnetic waves in empty space (3×10^8 meters per second) gives the virtual height.

13-4. Maximum Usable Frequencies (M-U-F).- It will be noted that the preceding discussion is based on vertical incidence transmission, that is, the wave is sent practically vertically upwards and the reflection, if any, is received near the sending point. The results of vertical-incidence transmission can be used to determine "maximum usable frequencies." A maximum usable frequency (m-u-f) is the highest frequency which can be used at any given time and between any given pair of points on the earth's surface, for radio transmission by reflection from the regular layers of the ionosphere. (Transmission at higher frequencies is sometimes possible by sporadic and scattered reflections.) It was shown in Ch. 12, that for a plane wave passing through one dielectric medium and striking the plane surface separating that medium from a second dielectric medium, the angle of incidence φ is equal to the angle of reflection, and the angle φ_t of the normal to the wave transmitted through the second medium is related to φ by

$$\frac{\sin \varphi}{\sin \varphi_T} = \frac{\beta_2}{\beta_1} = \frac{\lambda_1}{\lambda_2} = \frac{c_1}{c_2} = \sqrt{\frac{\varepsilon_1}{\varepsilon_2}} = \eta \tag{13-12}$$

where $\beta = 2\pi/\lambda$, λ is wavelength, c is phase velocity, ε is dielectric constant, 1 refers to the first medium and 2 to the second medium. For total reflection (no transmission in medium 2) φ_T must be 90°, or

$$\sin^2 \varphi = \eta^2 = 1 - \frac{f_c^2}{f^2} \tag{13-13}$$

For all values of φ equal to or greater than that given by (13-13), there will be total reflection, hence

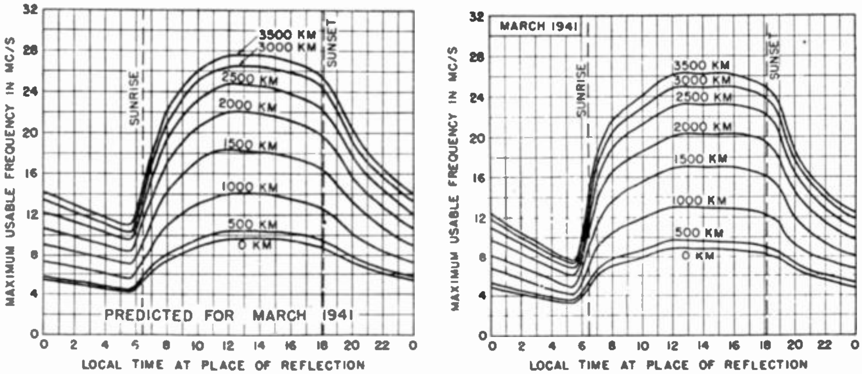
$$f = f_c \sec \varphi \tag{13-14}$$

gives the maximum frequency for total reflection. For smaller values of the equivalent dielectric constant of the medium is negative, and the practical result is that the wave passes through the layer. See Fig. 13-7. Equation (13-14) is the so-called "secant law," which states that the greatest frequency f usable at an angle of incidence φ is $f_c \sec \varphi$, where f_c is the critical frequency for vertical incidence.

If φ is given, f is immediately determined. But if the distance d between two points on the earth's surface (here assumed flat) is given, the determination of the m-u-f requires a knowledge of the virtual height of the ionosphere layer at the point of reflection. Since the virtual height z_v is a function of frequency it is necessary to solve simultaneously

$$z_v = F(f) \text{ and } f \sec \varphi = f_c \tag{13-15}$$

Along with curves of the type shown in Fig. 13-2, it is possible to obtain, by using numerous frequencies, a curve[9] of virtual height z_v vs frequency f. With such a curve and the secant law, a graphical solution of the two conditions can be obtained, and the m-u-f thus determined.[10] Figure 13-3b shows



(a) (b)

FIG. 13-3. a) Prediction by the National Bureau of Standards of maximum usable frequencies for dependable radio transmission via the regular ionosphere layers on undisturbed days of March 1941; published in the Proc. I.R.E., 28, No. 12, Dec. 1940 (the curves are for different distances of transmission as indicated).

b) Actual maximum usable frequencies for March 1941, as measured at the National Bureau of Standards; published in the Proc. I.R.E., 29, No. 3, March 1941.

9. By use of a continuously varying frequency, the entire curve can be obtained at one time. Some experimental results are reproduced by K.K. Darrow, Analysis of the Ionosphere, Bell Sys. Tech. Jour., 19, No. 4, July (1940).

10. Smith, Newbern, Extension of Normal-Incidence Ionosphere Measurements to to Oblique-Incidence Radio Transmission, Jour. of Research of the Nat. Bur. of Standards, 19, No. 1, July (1937); The Relation of Radio Sky-Wave Transmission to Ionosphere Measurements, Proc. I.R.E., 27, No. 5, May (1939).

m-u-f for various hours of an average undisturbed day in March 1941, for various distances d. It is possible by a study of the numerous factors affecting transmission to make in advance reasonable predictions of m-u-f. Figure 13-3a shows such a prediction made by the National Bureau of Standards for March 1931, and Fig. 13-3b shows the actual results of measurements made during the month by that institution.

13-5. Transmission in Various Frequency Bands.- Returning to the group of frequency bands listed in Sec. 13-2, Band I (50-550 kc) is the long-wave band below the broadcast frequencies. The ground wave is important for the shorter transmission distances (less than 1000 miles), and is stable and reliable. The sky wave is important for very long distances (some transatlantic telegraph channels are operated near the longest wavelength in this band). In general, the long waves obey the Austin-Cohen formula[11] for decrease in field intensity with distance. However, theoretical analysis has now reached a stage where the ground wave (direct, ground-reflected, and surface waves) can be computed without great difficulty. Norton[12] has recently brought together and systematized the work of many investigators, but the analysis which determines and unites vectorially (1) the direct wave, (2) the ground-reflected wave, and (3) the surface or guided wave, allowing for finite conductivity and dielectric constant of the earth, is too extensive to be reproduced here. Figure 13-4 reproduces an example given by Norton. The "inverse distance" curve is the result obtained in Ch. 12. The "plane earth" curve is that obtained by assuming the earth flat, and although earth conductivity and dielectric constant have been taken into account, it is seen that a "plane earth" introduces virtually no attenuation--in this case. When, however, the curvature of the earth is considered, the lowest curve results.

Band II, the broadcast band, has good and reliable service within 50 to 100 miles of the antenna; the distance depends on transmitter power as well as other factors. The ground wave accounts for daytime transmission. At night, transmission may extend much further because of the sky wave (see Fig. 13-8),

11. Mimno, loc. cit.; the Austin-Cohen formula was one of the first experimentally determined formulas for decrease of intensity with distance, and is often cited.

12. Norton, K.A., The Calculation of Ground-Wave Field Intensity over a Finitely-Conducting Spherical Earth, Proc. I.R.E., 29, No. 12, Dec. (1941).

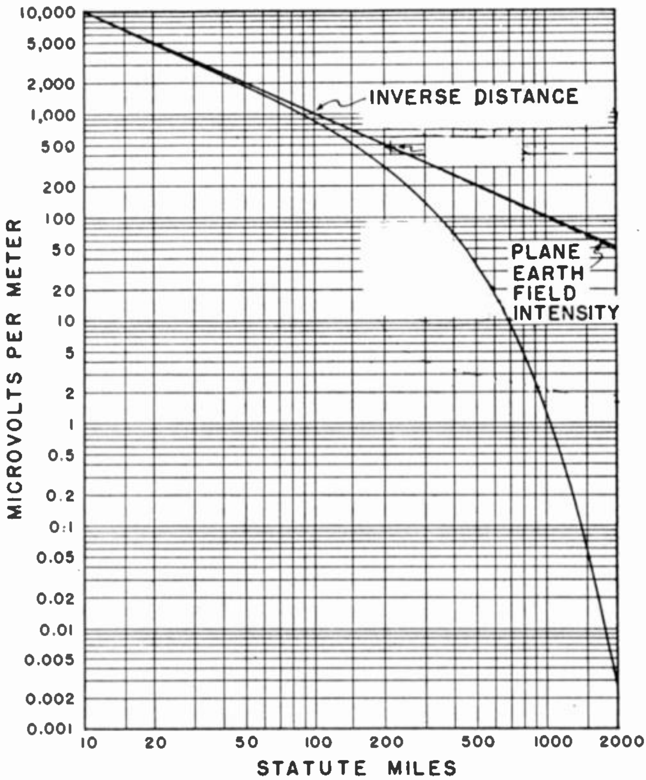


FIG. 13-4. Variation of rms electric field strength of ground wave with distance under following conditions: transmitting and receiving antennas linear and vertical (field is vertically polarized), and near the ground; frequency 500 kc; ground conductivity 5×10^{-11} emu, dielectric constant relative to air 80 (the ground constants are for sea water); the transmitting antenna produces in the equatorial plane at a distance of 1 mile a field strength of 50 millivolts per meter in free space or 100 millivolts per meter over a perfectly conducting earth. The inverse distance curve is thus 100/d vs d, where d is distance from the transmitter on the surface of the earth. The dielectric constant of the atmosphere is assumed to decrease linearly with altitude. From Norton, Proc. I.R.E., 29, No. 12, Dec. 1941. Other results given in this article enable curves such as those in the figure to be determined relatively easily.

which in the daytime is largely absorbed rather than reflected by ionosphere layers (D and E). The ground-wave field intensity, allowing both for refraction resulting from change in the dielectric constant with altitude in the lower atmosphere, and for ground conductivity and dielectric constant, can be

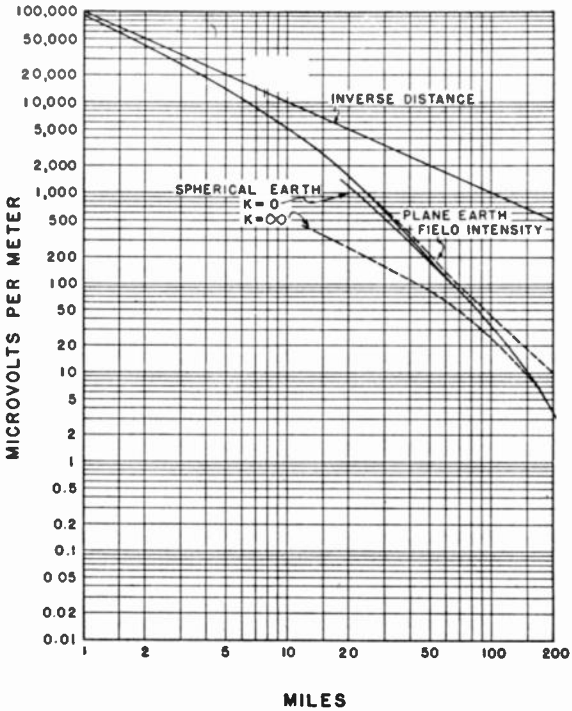


FIG. 13-5. Similar to Fig. 13-4, except for transmission at 1120 kc over land for which the conductivity has been assumed 9 x 10^{-14} emu and the relative dielectric constant 21. From Norton, loc. cit., note that in contrast to Fig. 13-4, the "plane earth" field intensity is not near the field intensity computed on the simple inverse distance basis.

determined from Norton's summary as for Band I. Figure 13-5 offers an illustration for a frequency falling in Band II.

Fading--"the variation in radio field intensity resulting from changes in the transmission medium"--is illustrated in Fig. 13-6. At A in the figure, the ground wave and sky wave are assumed to be received; if now through some disturbance, one of these changes, the receiver at A immediately notices this change. If the change is such as to throw the two waves out of phase temporarily, the received signal may be greatly reduced or may disappear. If this occurs for components of certain frequencies and not for others in the frequency band of the received wave, selective fading results.

Band III (1.5-5 Mc) of the group of frequency bands set out in Sec. 13-2 forms to a certain extent a transitional region from the broadcase band (II) to Band IV. Usually, antenna heights (not lengths--height is that of the center of the

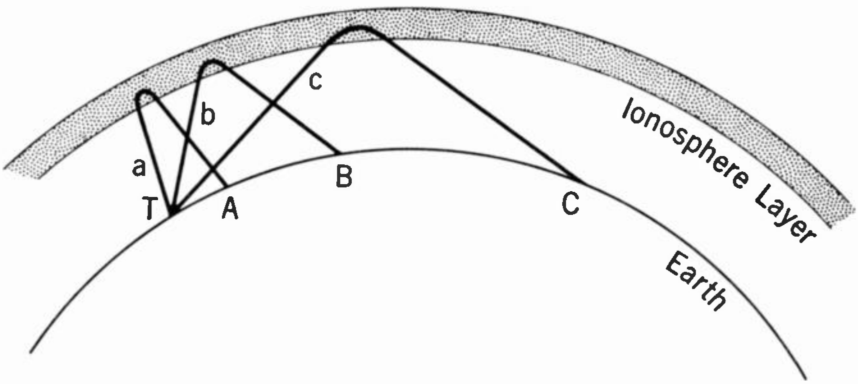


FIG. 13-6. Illustration of fading; if the ground wave at A has been reduced to a value comparable to that of the sky wave then fading will occur unless the ground-wave and sky-wave components of the signal remain in phase or at least in the same relative phase. If components of some frequencies remain in phase, others not, selective fading occurs.

antenna) are so low that the surface wave is the only important part of the ground wave. The sky wave becomes of increasing importance, particularly at the higher frequency end of the band. The low-frequency end is characterized by somewhat erratic properties.

Band IV (5-30 Mc) stands out because the ground wave is usually of no importance. This is the approximate band of frequencies which appears on the maximum-usable-frequency curves illustrated in Fig. 13-3, and is extremely useful for long-range transmission. Figure 13-7 illustrates skip distance and the silent zone region. Figure 13-8 illustrates a "double hop" in reaching the point B. The number of "hops" in a transmission is often called the order of reflection.

Because of the essential connection between Band IV and the ionosphere, some definitions are listed below. It should be understood that these and other concepts discussed or illustrated in the Band IV description are applicable in other bands.

"Sporadic reflections from a layer of the ionosphere are sharply defined reflections of substantial intensity from the layer at frequencies greater than the critical frequency of the layer. The intensity of the sporadic reflections generally decreases with increasing frequency. They are variable in respect to time of occurrence, geographic distribution, and range of frequencies in which observed. (Sporadic reflections are sometimes called abnormal reflections.)

"Scattered reflections from a region of the ionosphere are reflections composed of many components of different virtual

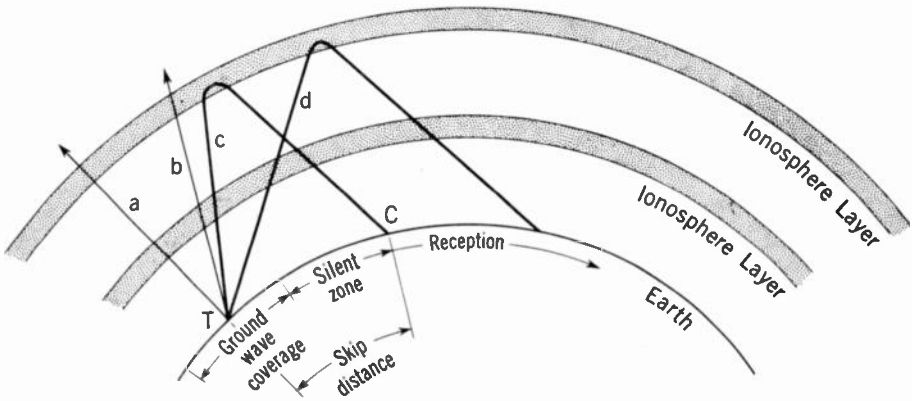


FIG. 13-7. Illustration of skip distance and the silent or skip zone. Path c is at maximum angle for signal to be reflected; paths a and b pass through ionosphere; path d is a path at less than the critical angle of path c. The ground nave is assumed to disappear outside the limits shown. No signal reaches the silent or skip zone, but beyond C reception is obtained. T is the transmitter.

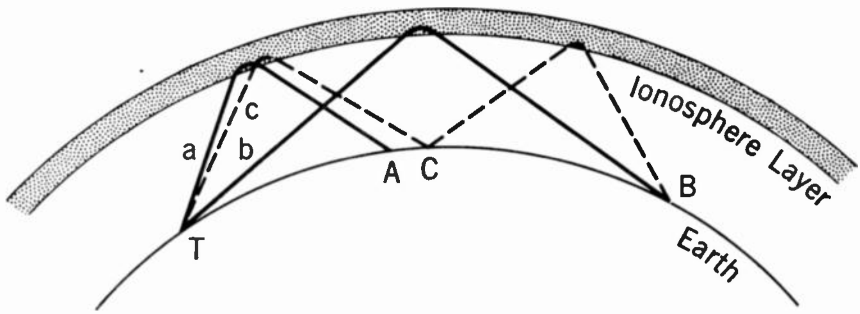


FIG. 13-8. Transmission by reflection from an ionosphere layer. Path a and b from the transmitter T proceed directly to A and B respectively; the wave following path c strikes the earth at C, is reflected to the ionosphere and back to the earth reaching B by a "double hop."

heights or times of travel which interfere and cause fluctuation or fluttering reception. They are quite variable in respect to time of occurrence, intensity, range of frequencies in which observed, etc.

"Zigzag reflections from a layer of the ionosphere are high-order multiple reflections which may be of abnormal intensity. They are produced by the return of waves which travel by multihop ionospheric reflection and finally turn back toward their starting point by repeated reflections from a slightly curved or sloping portion of an ionospheric layer.

"An ionospheric storm is a period of disturbance in the ionosphere in which there are great anomalies of penetration frequencies, virtual heights, and absorption. The phenomena are many and complex. They usually include: turbulence of the ionosphere and poorly defined penetration frequencies, especially at night; great virtual heights and low penetration frequencies, especially of the F and F_2 layers; great absorption, especially at night at broadcast frequencies. An ionospheric storm usually lasts a substantial part of a day or more than one day.

"A sudden ionospheric disturbance is a sudden increase of ionization density in low parts of the ionosphere, caused by a bright solar chromospheric eruption. The sudden ionospheric disturbance gives rise to a sudden increase of absorption in radio waves propagated through the low parts of the ionosphere, and sometimes to simultaneous anomalies in terrestrial magnetism, earth currents, etc. The change takes place within one or a few minutes, and conditions usually return to normal within one or a few hours.

"A radio fadeout is a cessation or near-cessation of propagation of radio waves through the parts of the ionosphere affected by a sudden ionospheric disturbance."

The vertical and horizontal angles of departure and of arrival are angles specifying the direction in which a wave leaves a transmitting antenna and the direction in which a wave approaches the receiving antenna.

In Band V (30-300 Mc), there is often transmission by an ionosphere layer, particularly at the lower frequencies, but this transmission is on the whole so unreliable as to be almost useless for practical purposes. In addition, the ground wave cannot be relied on at even medium distances, yet creates interference well beyond the dependable service area. The interference region beyond the service limit range is apparently the result of varying conditions in the troposphere. Substantially, this band is practical for line-of-sight transmission; the ground wave is the only important one (sky wave inconsequential except for sporadic results), and height of transmitting and receiving antennas becomes important.

Figure 13-9 is from Norton, and shows the decrease in field strength with distance under a given set of conditions for a transmission frequency of 46 Mc, which is in Band V. The field intensity scale is not immediately comparable with that of Figs. 13-5 and 13-6 because the transmitting antenna has been taken as a half-wave linear radiator (half-wave dipole) producing in the equatorial plane at a distance of 1 mile a field

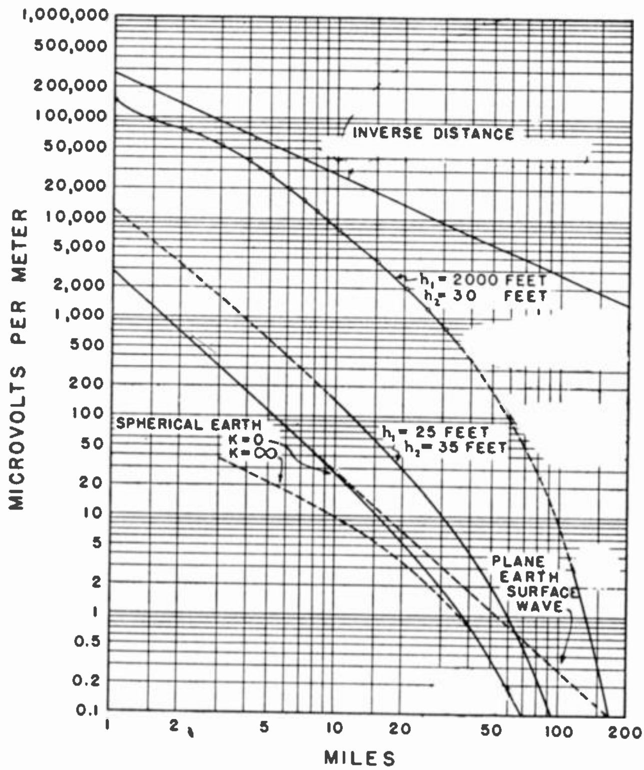


FIG. 13-9. Decrease of rms value of the ground-wave electric field strength (vertical polarization) with distance under the following conditions: frequency 46 Mc, conductivity of ground 5 x 10^{-14} emu, relative dielectric constant 15 (these values are for land); transmitting antenna vertical, center at height h_1; receiving antenna vertical, center at height h_2; the transmitting antenna produces in each case at 1 mile 137.6 mv/m in empty space or 275.2 mv/m if attached to a perfectly conducting earth. The inverse distance curve is 275.2/d where d is distance from the transmitting antenna. From Norton, loc. cit.

strength of 137.6 millivolts per meter in empty space, or 275.2 mv/m if connected to a perfectly conducting earth. For the antenna under either of these conditions, the power radiated is 1 kw. In this case all three parts of the ground wave are of importance.

Band VI (above 300 Mc) has line-of-sight transmission and little more. It follows Band V characteristics in some respects, but the surface-wave part of the ground wave is almost always negligible, so that the received signal is the vector sum of the direct and ground-reflected waves.

Chapter 14

HOLLOW WAVE GUIDES

It has been observed (Sec. 11-10) that the resistance of a transmission line at ultra-high frequencies is inversely proportional to the perimeter of a cross section of the conductor. The loss in a coaxial line is less than that in a parallel pair of wires of comparable overall size, because the tube has an inside perimeter much larger than that of the wires. There is also the advantage, for the coaxial line, that no appreciable radiation or interference field escapes through the outer metal tube as long as the frequency is high because skin effect keeps the current near the inner surface of the outer conductor.

The outer conductor of a coaxial line is usually at least 3 or 4 times as great in diameter as the center conductor, and so contributes only a small fraction of the loss of the entire line. It may be surmised therefore that if it is found possible to transmit waves through a simple hollow tube having no central conductor, then the losses will be very small, since they will be produced only in the large-perimeter low-resistance inside "skin" of the tube. Dielectric loss is neglected in this argument, but this assumption is more justifiable for a hollow tube than for a coaxial line, because in a hollow tube no solid dielectric supports are required to maintain the position of a center conductor.

It may appear astonishing to find that power may be transmitted inside a single hollow conductor when we compare this phenomenon with the ordinary transmission of power requiring two or more conductors. On the other hand, when approached from the viewpoint of an electromagnetic wave in a dielectric, it appears quite reasonable, since ordinary radio waves may be guided by the earth's surface or by the Kennelly-Heaviside layer (ionosphere) without a second conductor.

Hollow wave guides are similar to transmission lines in that their primary use is to "guide" or conduct energy from one point to another. They should be considered as simply one type of transmission line, just as a pair of parallel wires, a submarine cable, a coaxial line, etc. are other types. But hollow wave guides cannot conveniently be treated by ordinary transmission line theory as developed in Ch. 11. Instead, it is desirable to use the same basis of treatment as that of radiation (Ch. 12), and for this reason they appear at this position in the text. In a limited sense they may be considered as a transition case between unguided radiation,[1] discussed in Ch. 12, and ordinary transmission lines.

14-1. The Equations of Wave-Guide Propagation.- To understand anything more than the mere fact that under certain conditions energy may be transmitted by a hollow wave guide of conducting material, it is necessary to develop a theory which is based on the Maxwell equations obtained in Ch. 12. It was shown there that in a homogeneous isotropic non-conducting medium free of electric charge, such as the space inside a copper pipe, the following must hold[2]

$$\nabla^2 \underline{E} = \frac{1}{c^2}\frac{\partial^2 \underline{E}}{\partial t^2} \quad \text{and} \quad \nabla \cdot \underline{E} = 0 \tag{14-1}$$

$$\nabla^2 \underline{H} = \frac{1}{c^2}\frac{\partial^2 \underline{H}}{\partial^2 t^2} \quad \text{and} \quad \nabla \cdot \underline{H} = 0 \tag{14-2}$$

These are four equations relating the vector electric field intensity $\underline{E}$, the vector magnetic intensity $\underline{H}$, dielectric constant ε, permeability μ ($c^2 = 1/\mu\varepsilon$), time t, and position which may be expressed in terms of the rectangular coordinates x, y, z, of a point, or in terms of other coordinates such as the cylindrical coordinates of a point.

Consider now a hollow cylindrical[3] conducting tube of constant but arbitrary cross section and let the X-axis be the axis of the tube. Assume the walls of the hollow pipe are

---

1. Antenna radiation into space is usually considered unguided except when influenced by the earth or the ionosphere. "Guiding" as used here means guiding during transmission, and does not refer to directional effects introduced at the transmitting point by antenna arrays, etc.
2. The first part of Ch. 12 is devoted to the derivation of these equations and others needed in this chapter, and to expressing them in vector form for simplicity. Units used here are also given in Ch. 12.

   In the equations (14-1),

$$\nabla^2 \underline{E} = \frac{1}{c^2}\frac{\partial^2 \underline{E}}{\partial t^2}$$

   really represents a group of three separate equations for the components of $\underline{E}$; for rectangular components E_x, E_y, E_z, each component equation is of the form

$$\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 E_x}{\partial t^2} \tag{14-1a}$$

   The equation $\nabla \cdot \underline{E} = 0$ is, written out,

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0. \tag{14-1b}$$

   Equations (14-2) give similar results for $\underline{H}$.
3. Cylindrical does not mean circular cross section necessarily.

perfect conductors.

Can electromagnetic energy be transmitted through such a tube? If fields $\underline{E}$ and $\underline{H}$ can be found which satisfy (14-1) and in addition satisfy the condition which must be met at the inner surface of the pipe, the answer is yes; otherwise it is no. Since the pipe is assumed a perfect conductor, the condition at the inner surface of the pipe is simply that the component of the electric field tangent to the surface must be zero at every point of the surface at all times. The material which follows is thus an attempt to answer the query: Are there $\underline{E}$ and $\underline{H}$ fields which satisfy (14-1) and (14-2) and the surface condition?

Since, as discussed in Ch. 1, any periodic function (with restrictions inconsequential practically) may be expressed in a Fourier series, and a non-periodic function by a Fourier integral, we may consider without essential loss of generality a solution for $\underline{E}$ in which variations with respect to x and time are both sinusoidal. Accordingly, we assume

$$\underline{E} = \underline{F}(y,z) \cos (\omega t - \beta x) \tag{14-3}$$

where $\underline{F}(y,z)$ is any function of y and z, but not of x or t. Equation (14-3) will represent a wave traveling in the X direction[4] when β and ω are real quantities. From (14-3) it follows that

$$\frac{\partial^2 \underline{E}}{\partial x^2} = -\beta^2 \underline{E} \quad \text{and} \quad \frac{\partial^2 \underline{E}}{\partial t^2} = -\omega^2 \underline{E}. \tag{14-4}$$

Here ω is the angular velocity $2\pi f$, and β is $2\pi/\lambda_g$, where λ_g is the wavelength[5] along the X-axis of the wave which is moving in the guide. Making these substitutions in the wave equation (14-1) there results[6]

$$\frac{\partial^2 \underline{E}}{\partial y^2} + \frac{\partial^2 \underline{E}}{\partial z^2} = -k^2 \underline{E} \tag{14-5}$$

4. See Ch. 11, where traveling waves are discussed and a graph of the progression of a damped traveling wave is given. Note that the wave will travel in the positive X direction, and $\underline{F}(y,z)$ will tell how the magnitude varies over any cross section normal to X.

5. It has been pointed out in Ch. 11 and Ch. 12 that a wavelength is the amount by which x in (14-3) must change, t remaining fixed, so that the argument of the sign changes by 2π. Hence $\beta = 2\pi/\lambda_g$. It is important to note here the subscript g on λ, since it will be shown shortly that the wavelength in the guide is different from the wavelength in an unbounded medium.

6. The vector equation written in (14-5) is equivalent to three scalar equations which in rectangular component form are

$$\frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} = -k^2 E_x \tag{14-5a}$$

where the constant $k^2 = \omega^2/c^2 - \beta^2$.

The components of magnetic field intensity are related to those of electric field intensity by the third Maxwell law[7]

$$-\mu \frac{\partial \underline{\underline{H}}}{\partial t} = \nabla \times \underline{\underline{E}} \tag{14-6}$$

The boundary condition, that at the surface of a perfect conductor the component of $\underline{\underline{E}}$ tangential to the surface shall vanish, requires that at the boundary or surface

$$E_x = 0 \tag{14-10}$$

and

$$E_y \frac{dy}{ds} + E_z \frac{dz}{ds} = 0 \tag{14-11}$$

where ds is an infinitesimal arc of the curve in which the YZ-plane intersects the inside surface of the hollow conducting cylinder.[8]

14-2. Rectangular Section.- The simplest solution results when the transverse cross section of the hollow tube is rectangular. Let the edges of the rectangle be a and b units in length parallel respectively to the Y- and Z-axes, with the X-axis coinciding with one longitudinal edge. See Fig. 14-1.

(footnote continued)

$$\frac{\partial^2 E_y}{\partial y^2} + \frac{\partial^2 E_y}{\partial z^2} = -k^2 E_y \tag{14-5b}$$

$$\frac{\partial^2 E_z}{\partial y^2} + \frac{\partial^2 E_z}{\partial z^2} = -k^2 E_z \tag{14-5c}$$

7. The equation is discussed in Ch. 12. The vector form given in (14-6) represents the following three equations in rectangular coordinates

$$-\mu \frac{\partial H_x}{\partial t} = \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \tag{14-7}$$

$$-\mu \frac{\partial H_y}{\partial t} = \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \tag{14-8}$$

$$-\mu \frac{\partial H_z}{\partial t} = \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \tag{14-9}$$

8. Since dy/ds and dz/ds are the direction cosines of ds, (14-11) states that the sum of the projections of E_y and E_z on ds must be zero. In other words, $E_x = 0$ assures that of the total component of $\underline{E}$ tangent to the surface, the part in the X direction must be zero, and (14-11) requires that the part perpendicular to X must be zero.

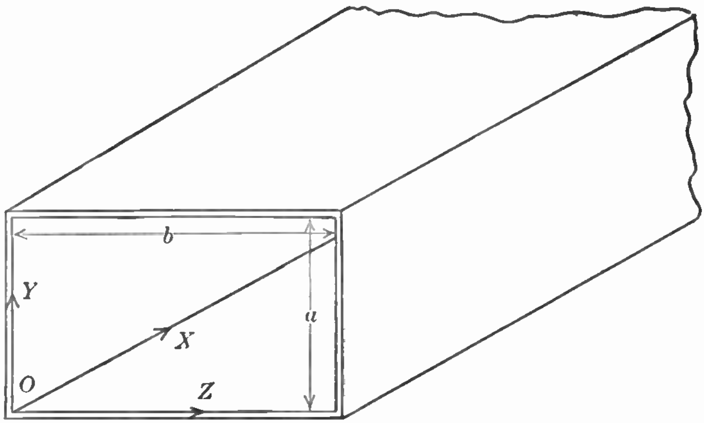


FIG. 14-1. Hollow tube of rectangular section, showing axes.

The four corners are located at $y = 0$ and a, $z = 0$ and b. Consider first the solution to the equation (14-7), namely

$$\frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} = -k^2 E_x,$$

subject to the boundary conditions stated. This equation and the boundary conditions are satisfied if

$$E_x = 0 \tag{14-12}$$

or if a field such as that shown by the solid lines of Figs. 14-3 and 14-5 exists, which indicates a possible solution

$$E_x = A \sin \frac{m\pi y}{a} \sin \frac{n\pi z}{b} \cos (\omega t - \beta x) \tag{14-13}$$

where m and n must be integers in order to make E_x zero at $y = a$ and $z = b$, and there is the further restriction, obtained by substituting (14-13) in (14-7), that

$$k^2 = \pi^2 \left(\frac{m^2}{a^2} + \frac{n^2}{b^2}\right). \tag{14-14}$$

It may be observed that solution (14-12) is merely the special case of (14-13) in which $m = 0$ or $n = 0$.

Unless both m and n are each at least unity, eq. (14-13) is identically zero and the field within the tube is zero everywhere, and when m and n are each just unity, it follows that for this wave, called the $TM_{1,1}$ wave.[9] Since k^2 has been defined as $\omega^2/c^2 - \beta^2$, it is seen that

$$\beta^2 = \frac{\omega^2}{c^2} - \frac{\pi^2}{a^2} - \frac{\pi^2}{b} \qquad (14\text{-}16)$$

and the lowest possible value of ω for β to be real ($\beta^2 > 0$) is

$$\omega^2 = c^2 \frac{\pi^2}{a^2} + \frac{\pi^2}{b^2} \qquad (14\text{-}17)$$

For this exact value of ω, the propagation phase constant β is zero, indicating an infinite velocity of phase propagation in the X or longitudinal direction.[10]

The value of the phase constant, β, is for any values of m and n

$$\beta = \frac{\omega}{c}\sqrt{1 - \frac{\pi^2 c^2}{\omega^2}\left(\frac{m^2}{a^2} + \frac{n^2}{b^2}\right)} \qquad (14\text{-}18)$$

The solutions for E_y and E_z from eq. (14-5b) and (14-5c) are of the same type, since the differential equations are identical in form, but must be subject to the boundary conditions, respectively, that $E_y = 0$ at $z = 0$ and $z = b$, and that $E_z = 0$ at $y = 0$ and $y = a$. The three components must also be related by the continuity condition $\nabla \cdot \underline{E} = 0$ expressed in component form in eq. (14-1b), which determines the relative sizes of the amplitude coefficients. We have therefore

$$E_x = A \sin\frac{m\pi y}{a} \sin\frac{n\pi z}{b} \cos(\omega t - \beta x); \qquad (14\text{-}19)$$

$$E_y = A \frac{\beta}{k^2}\frac{m\pi}{a} \cos\frac{m\pi y}{a} \sin\frac{n\pi z}{b} \sin(\omega t - \beta x); \qquad (14\text{-}20)$$

$$E_z = A \frac{\beta}{k^2}\frac{n\pi}{b} \sin\frac{m\pi y}{a} \cos\frac{n\pi z}{b} \sin(\omega t - \beta x). \qquad (14\text{-}21)$$

When the components of $\underline{E}$ have been found, the components of $\underline{H}$ are easily obtained by substitution in the third Maxwell

9. The TM stands for "transverse magnetic." It is shown below that $H_x = 0$, hence the magnetic field is transverse, that is, has no component in the direction of propagation (X) of the wave.

10. Phase velocity = λ/T where T is the period ($T = 2\pi/\omega$). The distinction between phase velocity and actual velocity of energy propagation was brought out in Ch. 11 in connection with transmission lines; the discussion applies equally well here.

Equation (14-18) indicates that β in the hollow wave guide is always less than ω/c. Since ω/c is $2\pi/\lambda$ where λ is the wavelength in an unlimited medium, λ_g is either greater than or at the limit equal to λ.

law as expressed in eq. (14-6). The H components as thus obtained are:

$$H_x = 0; \tag{14-22}$$

$$H_y = -A \frac{\varepsilon\omega}{k^2} \frac{n\pi}{b} \sin \frac{m\pi y}{a} \cos \frac{n\pi z}{b} \sin(\omega t - \beta x) \tag{14-23}$$

$$H_z = A \frac{\varepsilon\omega}{k^2} \frac{m\pi}{a} \cos \frac{m\pi y}{a} \sin \frac{n\pi z}{b} \sin(\omega t - \beta x) \tag{14-24}$$

Equations (14-22) to (14-24) can be rewritten

$$H_x = 0 \tag{10-25}$$

$$H_y = -\varepsilon \frac{\omega}{\beta} E_z \tag{10-26}$$

$$H_z = \varepsilon \frac{\omega}{\beta} E_y \tag{10-27}$$

which shows that the essential difference in the fields lies in the fact that E has a longitudinal component E_x; that is, a component in the direction of propagation down the hollow guide.

A plot[11] of the electric and magnetic field intensities is shown in Fig. 14-2 for the type of wave just derived, with m = 1 and n = 1. This wave is designated as the $TM_{1,1}$ wave, since it has a transverse component of H, the double subscripts indicating the values of m and n.

The phase velocity $v_{m,n}$ in an air-filled pipe a x b for the $TM_{m,n}$ wave is

$$v_{m,n} = \frac{\omega}{\beta_{m,n}} = \frac{c}{\sqrt{1 - (\omega_{m,n}/\omega)^2}} \tag{14-28}$$

where $\beta_{m,n}$ is the phase constant given by

$$\beta^2_{m,n} = \frac{\omega^2}{c^2} - \pi^2 \left(\frac{m^2}{a^2} + \frac{n^2}{b^2}\right), \tag{14-29}$$

11. The reader will find in the literature on hollow wave guides many plots of field lines. This is because (1) the theory is built up in terms of fields, hence the field distribution is of as much interest here as is the current distribution along an ordinary two-wire transmission line; (2) methods of supplying energy to wave guides and extracting energy from them depend to a considerable extent on field distribution; and (3) calculation of energy transmitted depends on the use of Poynting's vector (Ch. 12) which in turn requires a knowledge of the field distribution.

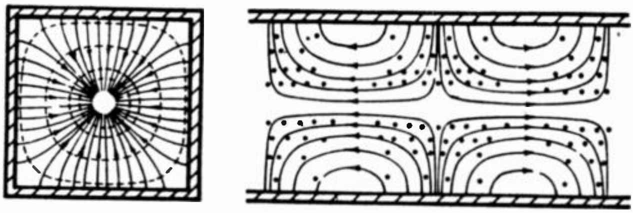


FIG. 14-2. The $TM_{1,1}$ wave. Solid lines are field lines of E, dotted lines of H. The configuration shown in the side view, which is a section parallel to the side and passing through the center of the tube, is to be pictured moving to the right at a speed equal to the phase velocity. In the side view, a solid dot is a magnetic field line coming out of paper, a circle is one going into paper. (Chu and Barrow.)

and $\omega_{m,n}$ is the cutoff angular frequency[12] for the same type of wave, equal to

$$\omega_{m,n} = c\,\pi\sqrt{\frac{m^2}{a^2} + \frac{n^2}{b^2}}\,. \tag{14-30}$$

The group[13] velocity $V_{m,n}$ is

$$V_{m,n} = \frac{\partial\omega}{\partial\beta_{m,n}} = \frac{c^2\beta_{m,n}}{\omega} = \frac{c^2}{v_{m,n}} \quad \text{or} \quad V_{m,n}v_{m,n} = c^2 \tag{14-32}$$

12. This is the most important single result obtained. At a frequency below that given by (14-30), β^2 becomes negative, see (14-29), and hence β is imaginary. The fields are then not propagated, but are attenuated rapidly. As the reader can show from (14-30), hollow wave guides are essentially u-h-f devices, since the frequency must be high before a field can be transmitted through a hollow wave guide of reasonable size.
13. Group velocity is introduced here because it is sometimes of more importance than phase velocity. Group velocity can be illustrated thus: Consider two waves of the same amplitude but slightly different frequency and wavelength

$$A \cos(\omega t - \beta x) \tag{14-31a}$$

and

$$A \cos[(\omega + \Delta\omega)t - (\beta + \Delta\beta)x]$$

By trigonometric manipulation the sum of these can be reduced to

$$2A \cos\frac{1}{2}(t\Delta\omega - x\Delta\beta)\cos(\omega t - \beta x) \tag{14-31b}$$

The phase velocity of the wave (14-31a) is ω/β, that is $\cos(\omega t - \beta x)$ has the same value at time t_2 and position x_2 as it had at time t_1 at position x_1 provided

$$\omega t_1 - \beta x_1 = \omega t_2 - \beta x_2$$

The wavelength λ_g in the guide is $\lambda_g = 2\pi/\beta$, and since β in the guide is always less than the same quantity in unlimited space (see 14-29), λ_g is always greater than λ, the wavelength in unlimited space.

Rectangular Section. Transverse Electric (TE) Waves. It has been shown in the preceding paragraphs that fields can be propagated down a hollow wave guide provided the frequency is great enough, and provided the fields have certain definite distributions. An entirely different type of wave may be transmitted through the rectangular tube. We have just considered the TM wave having a component of $\underline{E}$ in the axial (X) direction, but no $\underline{H}$ component. Conversely, there may exist a set of waves having a component of $\underline{H}$, but none of $\underline{E}$, in the axial or X direction. To solve for this other type of wave, called TE or transverse electric waves, set

$$E_x = 0 \tag{14-33}$$

and write, as in eq. (14-5b and c),

$$\frac{\partial^2 E_y}{\partial y^2} + \frac{\partial^2 E_y}{\partial z^2} = -k^2 E_y \tag{14-34}$$

and

$$\frac{\partial^2 E_z}{\partial y^2} + \frac{\partial^2 E_z}{\partial z^2} = -k^2 E_z .$$

The equation containing E_x in (14-5) has dropped out since E_x and all of its derivatives are assumed zero. The value of k^2 is

(footnote continued)

or $\dfrac{\text{distance interval}}{\text{time interval}} = \dfrac{x_2 - x_1}{t_2 - t_1} = \dfrac{\omega}{\beta}$ = velocity of phase propagation.

In (14-31b) the situation is not so simple, but if $2A \cos \frac{1}{2} (t\Delta\omega - x\Delta\beta)$ is taken to be the magnitude of the wave, then this magnitude, by an argument exactly like that above, moves with a velocity

$$\frac{\Delta\omega}{\Delta\beta} \text{ or, for small differences, } \frac{d\omega}{d\beta}$$

This is the quantity known as group velocity. If, now, a pulse containing many components of different frequencies travels down a transmission line of any kind (including in this group hollow wave guides), it is more likely to travel at the group velocity rather than the phase velocity. The latter is substantially a steady-state concept, and the fact that in the steady-state a wave, or any given phase of the wave, seems to travel at the phase velocity does not say that the effect of an arbitrary change at the input end is transmitted at this velocity. Only after a steady alternating state is reached will the apparent velocity be the phase velocity.

Note that, by (14-32), the phase velocity in the guide will always be greater than c (= 3×10^8 meters per second) when the group velocity is less than c (always true).

the same as previously, that is:

$$k^2 = \frac{\omega^2}{c^2} - \beta^2. \tag{14-35}$$

Obviously the solutions of (14-34) are again sine and cosine functions, since the second derivatives are proportional to the negatives of the functions. Since the boundary conditions require that

$$E_y \frac{dy}{ds} + E_z \frac{dz}{ds} = 0 \tag{14-36}$$

in order for the tangential component of E to vanish at the boundary, we must have in any YZ-plane

$$E_y \propto \cos \frac{m\pi y}{a} \sin \frac{n\pi z}{b}; \tag{14-37}$$

$$E_z \propto \sin \frac{m\pi y}{a} \cos \frac{n\pi z}{b}. \tag{14-38}$$

Thus the total E becomes equal to E_y (i.e., $E_z = 0$) at the edges where $z = 0$ and $z = b$, and $E = E_z$ (i.e., $E_y = 0$) at the edges where $y = 0$ and $y = a$.

From the equation of continuity, eq. (14-1b), simplified by the fact that $E_x = 0$,

$$\frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0, \tag{14-39}$$

the relative size of the coefficients in eq. (14-37) and (14-38) is determined. The coefficient of the expression for E_y must be in proportion to the coefficient of the E_z expression in the ratio of $n\pi/b$ to $-m\pi/a$. Any common arbitrary multiplier may be used, and we find it convenient to write, in terms of an arbitrary amplitude coefficient B,

$$E_x = 0; \tag{14-40}$$

$$E_y = -B \frac{\omega\mu}{k^2} \frac{n\pi}{b} \cos \frac{m\pi y}{a} \sin \frac{n\pi z}{b} \sin (\omega t - \beta x); \tag{14-41}$$

$$E_z = B \frac{\omega\mu}{k^2} \frac{m\pi}{a} \sin \frac{m\pi y}{a} \cos \frac{n\pi z}{b} \sin (\omega t - \beta x). \tag{14-42}$$

The expressions for the components of H are derived from these three expressions, just as before for the E wave, by substituting in the third Maxwell law as stated in eq. (14-7). The results are

$$H_x = -B \cos \frac{m\pi y}{a} \cos \frac{n\pi z}{b} \cos (\omega t - \beta x); \tag{14-43}$$

$$H_y = B \frac{\beta}{k^2} \frac{m\pi}{a} \sin \frac{m\pi y}{a} \cos \frac{n\pi z}{b} \sin (\omega t - \beta x); \tag{14-44}$$

$$H_z = B \frac{\beta}{k^2} \frac{n\pi}{b} \cos \frac{m\pi y}{a} \sin \frac{n\pi z}{b} \sin (\omega t - \beta x). \tag{14-45}$$

It is seen that
$$E_x = 0$$
$$E_y = -\mu \frac{\omega}{\beta} H_z \tag{14-46}$$
and
$$E_z = \mu \frac{\omega}{\beta} H_y \tag{14-47}$$

When both m and n are zero, all components must vanish, but either m or n may be zero provided the other is an integer. The $TE_{0,1}$ wave (i.e., m = 0, n = 1) has the lowest cutoff frequency of all which can be transmitted through a given tube, and for this $TE_{0,1}$ type we have:

$$E_x = E_z = H_y = 0 \quad \text{and} \tag{14-48}$$
$$E_y = B \frac{\mu\omega\pi}{k^2 b} \sin \frac{\pi z}{b} \sin (\omega t - \beta x); \tag{14-49}$$
$$H_x = - B \quad \cos \frac{\pi z}{b} \cos (\omega t - \beta x); \tag{14-50}$$
$$H_z = B \frac{\beta\pi}{k^2 b} \sin \frac{\pi z}{b} \sin (\omega t - \beta x). \tag{14-51}$$

A plot of the E and H field lines for the $TE_{0,1}$ wave is shown in Fig. 14-3. Plots of the $TE_{1,1}$, the $TE_{0,2}$, the $TE_{1,2}$ waves are also shown, in Figs. 14-4 to 14-6 inclusive.

14-3. Circular Section.- For tubes of circular section it is convenient to use cylindrical coordinates. Let the X-axis be coincident with the axis of the tube, and ρ and θ be respectively the radius vector and angle to any point in a plane normal to the X-axis (Fig. 14-7).

The wave eq. (14-1) in cylindrical coordinates, corresponding to eq. (14-1a) in rectangular coordinates, may be derived analytically directly from the latter, or else it may be arrived at by setting up the fundamental electromagnetic relations initially in cylindrical coordinates. In either event we have the wave eq. (14-1) in terms of the three electric field intensity components E_x, E_ρ, and E_θ:

$$\left.\begin{aligned}
\frac{\partial^2 E_x}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial E_x}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 E_x}{\partial \theta^2} + \frac{\partial^2 E_x}{\partial x^2} &= \frac{1}{c^2}\frac{\partial^2 E_x}{\partial t^2}; \\
\frac{\partial^2 E_\rho}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial E_\rho}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 E_\rho}{\partial \theta^2} + \frac{\partial^2 E_\rho}{\partial x^2} &= \frac{1}{c^2}\frac{\partial^2 E_\rho}{\partial t^2}; \\
\frac{\partial^2 E_\theta}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial E_\theta}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 E_\theta}{\partial \theta^2} + \frac{\partial^2 E_\theta}{\partial x^2} &= \frac{1}{c^2}\frac{\partial^2 E_\theta}{\partial t^2}.
\end{aligned}\right\} \tag{14-52}$$

Assuming sinusoidal variation with x and t as in (14-3), we set

$$\underline{E} = \underline{F}(\rho,\theta) \cos (\omega t - \beta x) \tag{14-53}$$

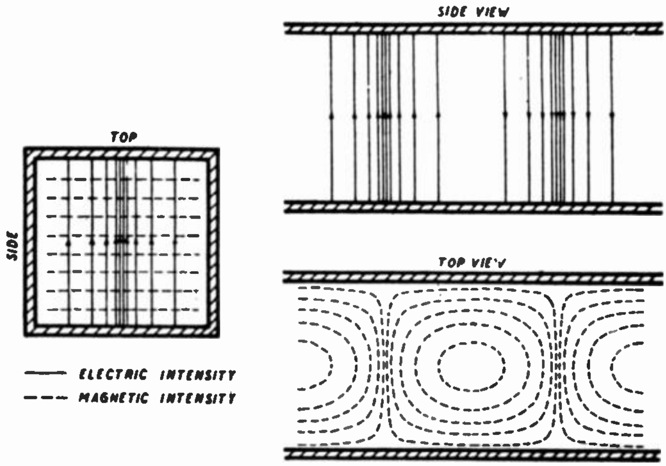


FIG. 14-3. The $TE_{0,1}$ wave.

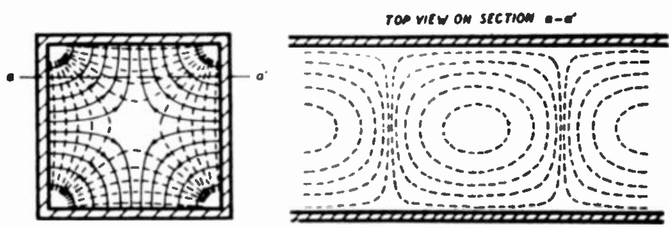


FIG. 14-4. The $TE_{1,1}$ wave.

Solid lines are field lines of $\underline{E}$, dotted lines of $\underline{H}$. The configuration shown in the side view is to be pictured moving to the right at a speed equal to the phase velocity (Chu and Barrow.)

whence

$$\frac{\partial^2 \underline{E}}{\partial x^2} = -\beta^2 \underline{E} \text{ and } \frac{\partial^2 \underline{E}}{\partial t^2} = -\omega^2 \underline{E} \tag{14-54}$$

as in (14-4). By these substitutions, eq. (14-52) reduces to

$$\left.\begin{aligned} \frac{\partial^2 E_x}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial E_x}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 E_x}{\partial \theta^2} &= -k^2 E_x; \\ \frac{\partial^2 E_\rho}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial E_\rho}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 E_\rho}{\partial \theta^2} &= -k^2 E_\rho; \\ \frac{\partial^2 E_\theta}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial E_\theta}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 E_\theta}{\partial \theta^2} &= -k^2 E_\theta. \end{aligned}\right\} \tag{14-55}$$

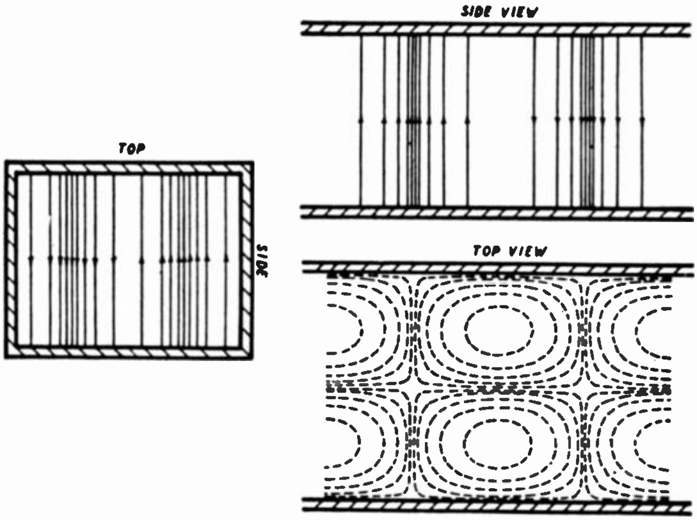


FIG. 14-5. The $TE_{0,2}$ wave.

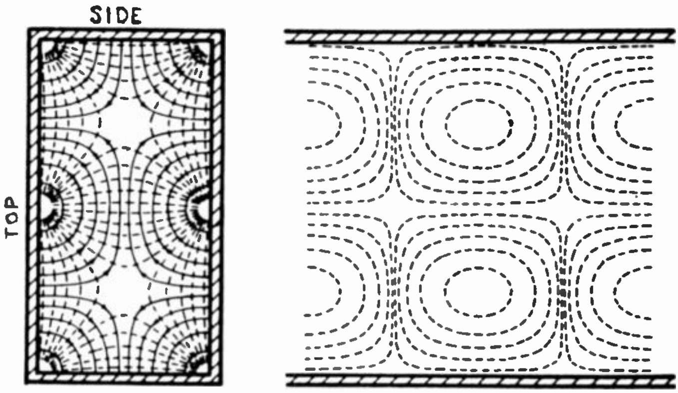


FIG. 14-6. The $TE_{1,2}$ wave.

Solid lines are field lines of $\underline{E}$, dotted lines of $\underline{H}$. The configuration shown in the side view is to be pictured moving to the right at a speed equal to the phase velocity. (Chu and Barrow.)

The boundary condition for E_x is that $E_x = 0$ at $\rho = a$, on the assumption of a perfectly conducting tube wall. To solve, we may assume that

$$E_x = P_0 + P_1 \cos \theta + P_2 \cos 2\theta + \ldots. \qquad (14\text{-}56)$$

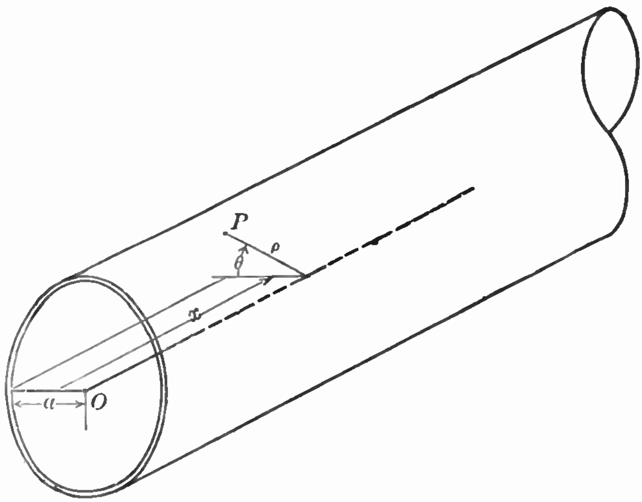


FIG. 14-7. Hollow tube of circular section, showing location of any point P by cylindrical coordinates x, ρ, θ.

where P_0, P_1, etc., are functions of ρ but not of θ. Substituting the general term $P_n \cos n\theta$ into the first of eq. (14-55) gives

$$\frac{\partial^2 P_n}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial P_n}{\partial \rho} + \left(k^2 - \frac{n^2}{\rho^2}\right)P_n = 0. \qquad (14\text{-}57)$$

This is Bessel's equation, having a solution $J_n(k\rho)$ which is everywhere finite, and a second solution with an infinite value at the origin, and variously designated as $G_n(k\rho)$, $K_n(k\rho)$ and $H_n(k\rho)$. For the present we must discard the second solution because no component of physical field within the tube can become infinite on the axis or elsewhere.

The first solution is derived readily by assuming a simple power series

$$P_n = a_0 + a_1\rho + a_2\rho^2 + \ldots \qquad (14\text{-}58)$$

and substituting into (14-57), then equating coefficients of like powers to zero. The resulting series is the Bessel function of the first kind and the nth order, $J_n(k\rho)$, equal to:

$$J_n(k\rho) = \frac{(k\rho)^2}{2^n n!}\left[1 - \frac{(k\rho)^2}{2^2 1!(n+1)} + \frac{(k\rho)^4}{2^4 2!(n+1)(n+2)} - \ldots\right] \qquad (14\text{-}59)$$

or in more compact form[14]

$$J_n(k\rho) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s!(n+s)!}\left(\frac{k\rho}{2}\right)^{n+2s}$$

which is the form in which the result was stated in connection with frequency modulation (Ch. 6).

The complete expression for the TM wave of the nth type, traveling in the positive X direction, is

$$E_x = A\, J_n(k\rho) \cos n\theta \cos(\omega t - \beta x). \qquad (14\text{-}60)$$

The E_ρ and E_θ components are given by the expressions:

$$E_\rho = A\, \frac{\beta}{k}\, J_n'\,(k\rho) \cos n\theta \sin(\omega t - \beta x) \qquad (14\text{-}60a)$$

and

$$E_\theta = -A\, \frac{\beta}{k^2}\, \frac{n}{\rho}\, J_n\,(k\rho) \sin n\theta \sin(\omega t - \beta x); \qquad (14\text{-}60b)$$

where J_n' indicates the derivative of J_n with respect to its argument.

The H components are obtained by using (14-65) to (14-67) and are

$$H_x = 0; \qquad (14\text{-}61)$$

$$H_\rho = A\varepsilon\, \frac{n\omega}{k^2\rho}\, J_n(k\rho) \sin n\theta \sin(\omega t - \beta x) \qquad (14\text{-}61a)$$

and

$$H_\theta = A\varepsilon\, \frac{\omega}{k}\, J_n'(k\rho) \cos n\theta \sin(\omega t - \beta x). \qquad (14\text{-}61b)$$

As in previous cases, it may be noted that some of the components of $\underline{E}$ and $\underline{H}$ are simply related to one another (compare (10-25) to (10-27) for TM waves in a rectangular guide):

$$H_x = 0$$

$$H_\rho = -\varepsilon\, \frac{\omega}{\beta}\, E_\theta \qquad (10\text{-}62)$$

$$H_\theta = \varepsilon\, \frac{\omega}{\beta}\, E_\rho \qquad (10\text{-}62a)$$

The values of k are restricted to those for which $J_n(ka) = 0$, in order to make E_x and $E_\theta = 0$ where $\rho = a$. A plot of the J_0 and of J_1 Bessel functions is shown in Fig. 14-8. The values of the argument at which these functions are zero may be read approximately from these curves. The values of these roots are, more precisely:

14. There are numerous texts on Bessel Functions, two of which are

A Treatise on Bessel Functions, etc., by Gray, Mathews, and Macrobert, Macmillan, 1931.

Bessel Functions for Engineers, by McLachlan, Oxford, 1934.

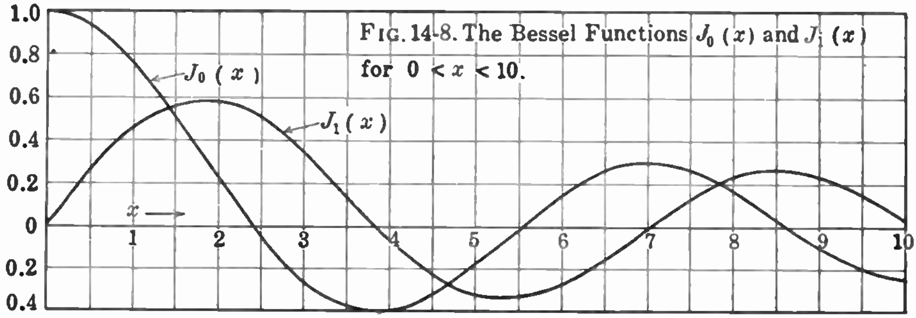


FIG. 14-8. The Bessel Functions $J_0(x)$ and $J_1(x)$ for $0 < x < 10$.

Roots of $J_0(x) = 0$	Roots of $J_1(x) = 0$	Roots of $J_2(x) = 0$
2.405	3.832	5.135
5.520	7.016	8.417
8.654	10.173	11.620

The simplest form of wave of the TM type and the one having the lowest cutoff frequency is that in which n = 0 and smallest root of the Bessel function is used; namely, 2.405.

Setting n equal to zero and giving k a value such that k a is equal to the first root of the zero-order Bessel function, 2.405, we have for the components of the $TM_{0,1}$ wave in a round tube:

$$E_x = A \quad J_0 \cos(\omega t - \beta x) \tag{14-63}$$

$$E_\rho = - A \frac{\beta}{u} J_1 \sin(\omega t - \beta x) \tag{14-63a}$$

$$E_\theta = H_x = H_\rho = 0 \tag{14-63b}$$

$$H_\theta = - A \frac{\omega}{u} J_1 \sin(\omega t - \beta x) \tag{14-63c}$$

in which $u \equiv 2.405/a$ and each J has an argument $u\rho$.

Note that these sets of equations satisfy the equation of continuity $\nabla \cdot \underline{E} = 0$, which in cylindrical coordinates is of the form

$$\frac{1}{\rho}\frac{\partial}{\partial \rho}(\rho E_\rho) + \frac{1}{\rho}\frac{\partial E_\theta}{\partial \theta} + \frac{\partial E_x}{\partial x} = 0 \tag{14-64}$$

and the $\underline{H}$ components are derived from those of $\underline{E}$ by the third Maxwell law, $\nabla \times \underline{E} = -\mu \frac{\partial H}{\partial t}$ which is here expressed:

$$\frac{1}{\rho}\frac{\partial E_x}{\partial \theta} - \frac{\partial E_\theta}{\partial x} = - \mu \frac{\partial H_\rho}{\partial t}; \tag{14-65}$$

$$\frac{\partial E_\rho}{\partial x} - \frac{\partial E_x}{\partial \rho} = - \mu \frac{\partial H_\theta}{\partial t}; \tag{14-66}$$

$$\frac{1}{\rho}\frac{\partial}{\partial \rho}(\rho E_\theta) - \frac{1}{\rho}\frac{\partial E_\rho}{\partial \theta} = - \mu \frac{\partial H_x}{\partial t}. \tag{14-67}$$

The wave just derived may be designated as the $TM_{0,1}$ wave, and waves of this general type as the $TM_{n,m}$ wave where n is the order of the Bessel function associated with the expression for the axial component of E, and m is the number of the Bessel function root employed. The fields are shown in Fig. 14-9.

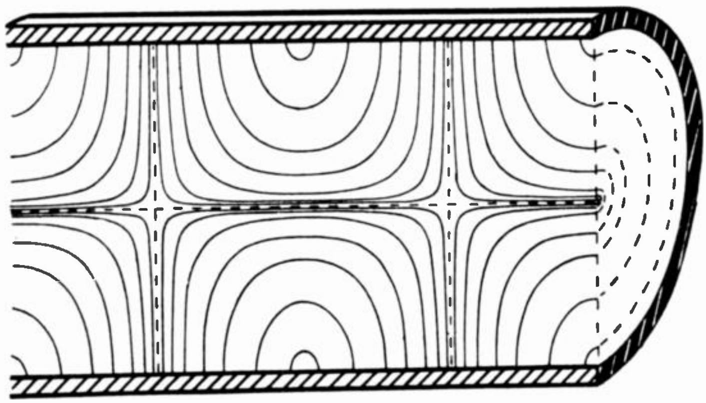


FIG. 14-9. The $TM_{0,1}$ wave in a circular tube. Solid lines of E, dotted lines of H. (Chu and Barrow.)

Circular Section. Transverse Electric (TE) Waves. Waves may exist having E_x equal to zero and H_x of finite value, analogous to the case of the tube with rectangular cross section. Using the same methods of analysis as those employed heretofore, we find:

$$E_x = 0; \tag{14-68}$$

$$E_\rho = B \frac{\omega n}{k^2 \rho} J_n(k\rho) \sin n\theta \sin (\omega t - \beta x) \tag{14-69}$$

$$E_\theta = B \frac{\omega}{k} J_n'(k\rho) \cos n\theta \sin (\omega t - \beta x) \tag{14-70}$$

$$H_x = B J_n(k\rho) \cos n\theta \cos (\omega t - \beta x) \tag{14-71}$$

$$H_\rho = B \frac{\beta}{k} J_n'(k\rho) \cos n\theta \sin (\omega t - \beta x) \tag{14-72}$$

$$H_\theta = - B \frac{\beta n}{k^2 \rho} J_n(k\rho) \sin n\theta \sin (\omega t - \beta x) \tag{14-73}$$

As in the other cases, some of the components of E and H are related:

$$E_x = 0$$

$$E_\rho = - \mu \frac{\omega}{\beta} H_\theta \tag{14-74}$$

$$E_\theta = \mu \frac{\omega}{\beta} H_\rho \tag{14-75}$$

Compare (10-46) and (10-47) for TE waves in a rectangular guide. The permissible values of k are those which make E_θ zero when $\rho = a$, that is those for which $J_n'(ka) = 0$. J_0' has its first zero at 3.832 and J_1' at 1.841. Thus the $TE_{1,1}$ mode has a higher cutoff value of λ than does the $TE_{0,1}$ mode.

For the $TE_{0,1}$ mode,

$$H_x = - B J_0 \cos (\omega t - \beta x) \tag{14-76}$$

$$H_\rho = B \frac{\beta}{u} J_1 \sin (\omega t - \beta x) \tag{14-77}$$

$$H_\theta = E_x = E_\rho = 0$$

$$E_\theta = \mu \frac{\omega}{\beta} H_\rho \tag{14-78}$$

in which $u \equiv 3.832/a$ and each J has an argument $u\rho$.

For this wave mode the attenuation may be made exceptionally small simply by increasing the frequency sufficiently; because the current in the conductor becomes less as the frequency is increased. This situation is radically different from the usual one where loss and attenuation increase indefinitely with increasing frequency.

For the $TE_{1,1}$ mode, which has the lowest cutoff frequency of all TM and TE waves in a pipe of circular cross section,

$$H_x = B \quad J_1 \cos \theta \cos (\omega t - \beta x) \tag{14-79}$$

$$H_\rho = B \frac{\beta}{2u} (J_0 - J_2) \cos \theta \sin (\omega t - \beta x) \tag{14-80}$$

$$H_\theta = - B \frac{\beta a}{w\rho} J_1 \sin \theta \sin (\omega t - \beta x) \tag{14-80a}$$

$$E_x = 0$$

$$E_\rho = - \mu \frac{\omega}{\beta} H_\theta \tag{14-81}$$

$$E_\theta = \mu \frac{\omega}{\beta} H_\rho \tag{14-81a}$$

in which $u \equiv 1.841/a$, $w \equiv 3.39/a$, and each J has an argument $u\rho$.

Figure 14-10 shows four field patterns for a circular hollow wave guide.

Dielectric Solid Wave Guides. It may be noted here that a solid dielectric "wire" has the property of guiding energy from one point to another. Such a solid dielectric wave guide has no metal associated with it. The theory is somewhat similar to that developed above, and the configurations of some possible fields in cylindrical solid dielectric wave guides of circular section resemble those shown in Fig. 14-10. However, there is no confinement of the electric field lines in the cylinder in the manner shown in the figure. The lines continue

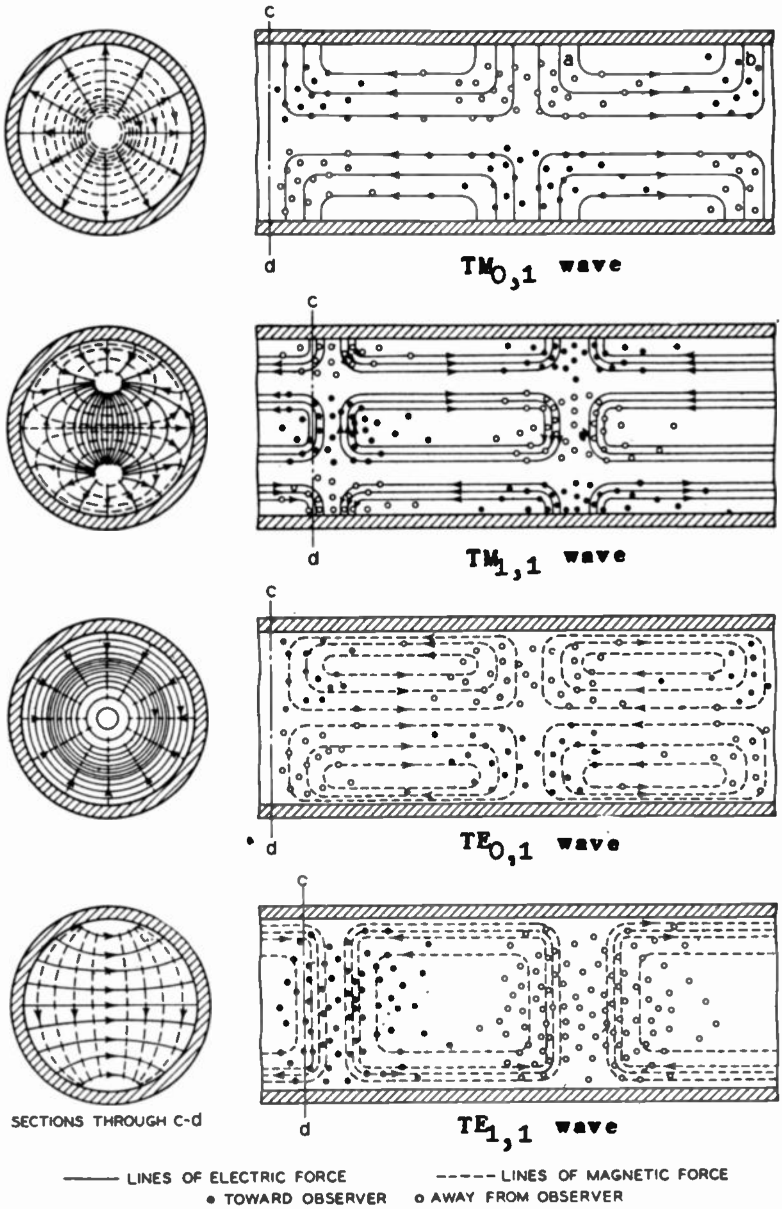


FIG. 14-10. Electric and magnetic lines of force in a hollow wave guide of circular section. Note particularly the $TE_{0,1}$ wave, for which the attenuation is very low (see Table 14-I). The waves move through the guide from left to right. Each right-hand picture is one wavelength long. (From Southworth, Proc. I.R.E., July 1937.)

into space, in a way determined by the appropriate boundary conditions.

14-4. Power Flow. The Poynting Vector.[15] - The flow of power in a hollow wave guide is calculated from the expressions for the E and H fields by the use of the Poynting vector P for density of power flow:

$$\underline{P} = \underline{E} \times \underline{H} \text{ watts per sq meter} \tag{14-82}$$

which has an axial component

$$P_x = E_y H_z - E_z H_y \text{ watts per sq meter} \tag{14-83}$$

in rectangular coordinates, or

$$P_x = E_\theta H_\rho - E_\rho H_\theta \text{ watts per sq. meter} \tag{14-84}$$

in cylindrical coordinates. The integral of P_x over a cross section of the dielectric is equal to the total power flow. Practically, we need consider only the simple modes of relatively low cutoff frequencies, namely the $TM_{1,1}$ and the $TE_{0,1}$, and perhaps the $TE_{1,1}$ modes. For the $TM_{m,n}$ mode in a rectangular pipe a x b meters in section,

$$P_x = A^2 \frac{\varepsilon\beta\omega\pi^2}{k^2}\left[\frac{m^2}{a^2}\left(\sin^2\frac{m\pi y}{a}\cos^2\frac{n\pi z}{b}\right) + \frac{n^2}{b^2}\left(\cos^2\frac{m\pi y}{a}\sin^2\frac{n\pi z}{b}\right)\right]\sin^2(\omega t - \beta x) \tag{14-85}$$

For the $TM_{m,n}$ mode in a square pipe a x a,

$$P_x = A^2 \frac{\varepsilon\beta\omega\pi^2}{k^2a^2}\left[\sin^2\frac{\pi y}{a}\cos^2\frac{\pi z}{a} + \cos^2\frac{\pi y}{a}\sin^2\frac{\pi z}{a}\right]\sin^2(\omega t - \beta x) \tag{14-86}$$

The average power, or average energy per unit time, passing in the X direction is obtained by first averaging P_x over a period [average of $\sin^2(\omega t - \beta x)$ is $\frac{1}{2}$] and then integrating the result over a cross section of the guide normal to X:

$$\mathrm{P}_x = \frac{1}{2}\int_0^a \int_0^b P_x \, dydz$$

$$= A^2 \frac{\varepsilon\beta\omega ab}{8} \text{ average watts (rectangle).} \tag{14-87}$$

If the pipe is square and the mode $TM_{1,1}$, this simplifies to:

$$\mathrm{P}_x = A^2 \frac{\varepsilon\beta\omega a^2}{8} \text{ average watts (square section).} \tag{14-88}$$

The constant A is equal to the maximum axial electric intensity

15. The Poynting vector is derived and discussed in Ch. 12.

E_x, which occurs at the center of the pipe. It is measured in volts per meter.

For the TE type of wave in a rectangular pipe,

$$P_x = B^2 \mu \frac{\beta \omega n^2 \pi^2}{8k^4 b^2} ab \quad \text{watts} \qquad (14\text{-}89)$$

which for the $TE_{0,1}$ mode becomes

$$P_x = B^2 \mu \frac{\beta \omega b^2}{\pi^2} ab \quad \text{watts} \qquad (14\text{-}90)$$

in which B is the maximum value of axial magnetic field, in ampere-turns per meter.

Many of the formulas for wave-guide transmission may be expressed in terms of cutoff frequency in a form which is independent of the shape of the cross section. It will be found convenient to collect the various formulas for reference and comparison.

14-5. **Formulas for Reference.**- In Table 14-I numerous formulas are collected. Note the important fact mentioned above that for the $TE_{0,1}$ wave in the round tube, alone of all the waves, the attenuation <u>decreases</u> as the frequency increases.

Concerning hollow wave guides, two basic questions are: Can they be used? and is it worth using them? The first question has been answered in the preceding work. The second question depends for an answer primarily on the attenuation. It is not feasible to enter here into the determination of attenuation for the various types of waves and guides. The calculation of losses and attenuation is often accomplished with sufficient precision by using the skin-effect resistance formula based on the equivalent depth of penetration, equal to $0.0662/\sqrt{f}$ meter for copper. The amount of charge at the terminals of the E lines is used to obtain a surface charge density, which multiplied by v gives a current density. The loss surface density is computed and integrated over the periphery. For the symmetrical E_0 and H_0 waves in round tubes the current density is independent of θ, but for the higher modes and in the rectangular pipes the current density varies.

The order of magnitude of the attenuation in copper tubes of size appropriate to 0.1 meter waves is 0.01 to 0.02 decibel per meter, far lower than values for coaxial cable or line of any other type when operated at this wavelength.

The wavelength λ_g in a guide, in terms of the free-space wave length λ and the cutoff wavelength λ_c, is

$$\lambda_g = \frac{\lambda \lambda_c}{\sqrt{\lambda_c^2 - \lambda^2}},$$

and λ in terms of λ_g and λ_c is

TABLE 14-I

Characteristics of Some Hollow Wave Guides

Type of Hollow Wave Guide	Type of Wave (Mode)	Cutoff Wavelength (λ_c)	Attenuation in db per meter (air core) $\eta = \frac{\lambda}{\lambda_c} = \frac{\omega_c}{\omega}$	Characteristic Impedance
Rectangular a x b (meters) b > a	$TE_{0,1}$	2b	$7930\sqrt{\frac{\rho}{\lambda}}\ \frac{b + 2a\,\eta^2}{ab\sqrt{1-\eta^2}}$	$\frac{465a}{b\sqrt{1-\eta^2}}$
	$TE_{1,1}$	$\frac{2ab}{\sqrt{a^2+b^2}}$		
	$TM_{1,1}$	$\frac{2ab}{\sqrt{a^2+b^2}}$		
Circular radius = a (meters)	$TE_{0,1}$	1.640 a	$11{,}450\sqrt{\frac{\rho}{\lambda}}\ \frac{\eta^2}{a\sqrt{1-\eta^2}}$	
	$TE_{1,1}$	3.412 a	$7930\sqrt{\frac{\rho}{\lambda}} \cdot \frac{0.418 + \eta^2}{a\sqrt{1-\eta^2}}$	$\frac{353}{\sqrt{1-\eta^2}}$
	$TM_{1,1}$	2.613 a	$11{,}450\sqrt{\frac{\rho}{\lambda}}\ \frac{1}{a\sqrt{1-\eta^2}}$	

Velocity of phase propagation: $v = \frac{3 \times 10^8}{\sqrt{1-\eta^2}}$ meters per second, all cases, air core.

Group velocity: $V = 9 \times 10^{16}/v$ meters per second.

For copper, specific resistance $\rho = 1.724 \times 10^{-8}$ ohm-meter.

$$\lambda_g = \frac{\lambda}{\sqrt{1-\eta^2}}$$ where λ_g is wavelength in guide, λ is wavelength in free space.

If the dielectric is not air or vacuum, the tabulated values of λ_c should be multiplied by $\sqrt{\varepsilon/\varepsilon_0}$ where ε is the dielectric constant of the core material and ε_0 that of air. The right-hand side of the equation for v above should be multiplied by $\sqrt{\varepsilon_0/\varepsilon}$.

$$\lambda = \frac{\lambda_g \lambda_c}{\sqrt{\lambda_g^2 + \lambda_c^2}}.$$

From the last equation it is possible to determine λ and hence the frequency from measurements by probes of wavelength in a guide whose cutoff wavelength is known.

For properly matching a guide to input and output impedances, a knowledge of the characteristic impedance of wave guides is helpful. The characteristic impedance for wave guides may be defined as the voltage-to-current ratio of a unidirectional wave, or as the quotient of power divided by current squared. There is a ratio of $\pi/4$ between values given by the first and second definitions, respectively.

The lowest characteristic impedance Z_K obtainable in a round pipe is 353 ohms. In a rectangular tube where a is less than b, Z_K may be given any value between 0 and 465 by varying the side a while the side b and the operating frequency remain fixed; because, the characteristic impedance is directly proportional to the narrow dimension a. By varying b as well, any value whatever for Z_K may be attained.

Southworth points out that the linear variation of Z_K with a indicates that the proper design of a junction or branch in a rectangular wave guide is to have the small dimension in the two branch lines equal each to one-half the small dimension a in the main line, with the longer dimension b remaining the same. This seems anomalous when compared with the branching of the ordinary two-wire transmission line, because there the two equal branches should have values of Z_K of twice that of the main line. In the wave guide, however, the transverse $\underline{E}$ field is halved at the junction, so that in a certain sense the two branches are in effect in series.

14-6. Excitation, Measurement of the Field Configuration, Use of Probes.- The excitation of a desired type of wave within a hollow tube may be accomplished by the insertion of a small antenna with an orientation designed to radiate a field approximating at least roughly the field of the desired wave type. Thus an antenna coincident with the axis would tend to set up TM waves, and a small loop normal to the axis would more successfully set up TE waves.[16] Other methods[17] are illustrated in Figs. 14-11 and 14-12.

---

16. Chu, L.J., and W.L. Barrow, Electromagnetic Waves in Hollow Metal Tubes of Rectangular Cross Section, Proc. I.R.E., 26, No. 12, Dec. (1938).

17. Southworth, G.C., Some Fundamental Experiments with Wave Guides, Proc. I.R.E., 25, No. 7, July (1937).

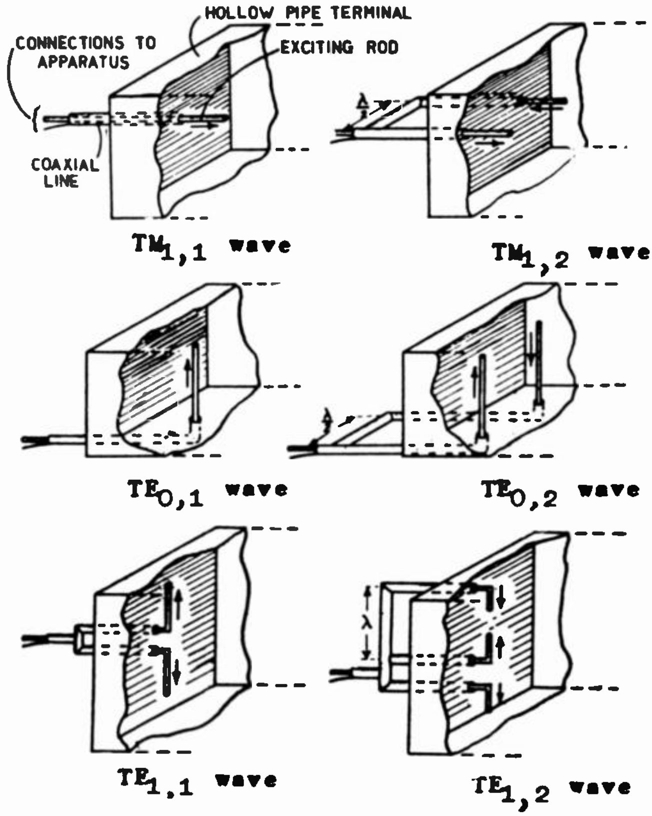


FIG. 14-11. Arrangements for setting up different types of waves in a rectangular guide. (Chu and Barrow.)

The configuration of the electric field lines may be demonstrated by the use of a probe with a crystal detector or rectifier. Two short pick-up wires or antennas are used to probe in the wave-guide field, and bring power to the crystal, from which connection is made to a microammeter. The probe should be short; for example, for 10-cm waves the total length of the detector with its antennas may be about an inch. The crystal, with its two projecting probe antennas, may be mounted at the end of a thin bakelite tube, through which the d-c leads may be brought. The direction of the electric field lines at any given point may be determined by holding the probe at that point and turning it until a maximum reading of the microammeter is obtained. The direction of the field will then be the same as that of the probe wires. The reading of the instrument will afford at least an approximate quantitive idea of relative

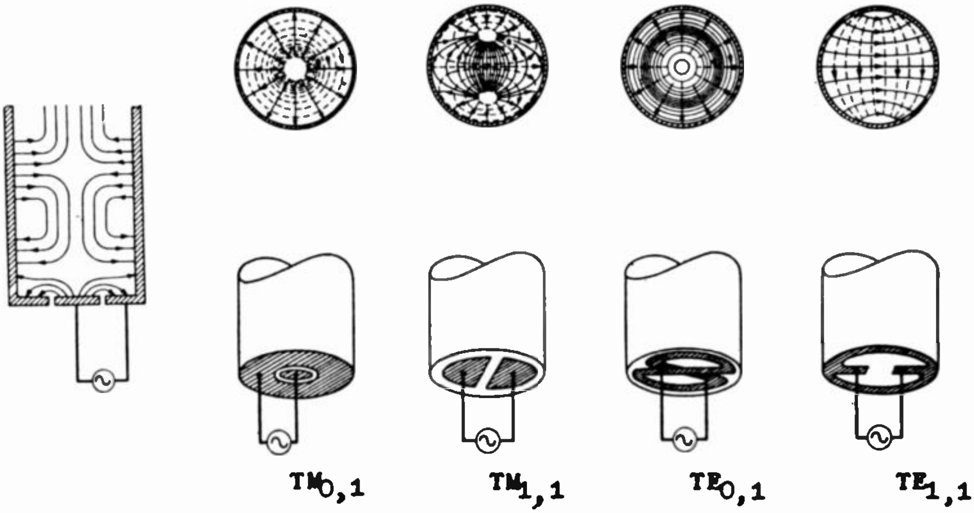


FIG. 14-12. Arrangements for setting up different types of waves (Southworth). The left-hand figure shows formation of the $TM_{0,1}$ wave.

field strengths, when the probe is placed in different positions and with different orientations.

Probes of this type are useful in many different ways. If the antenna system is less than a quarter wavelength, the response will be nearly linear even though the signal frequency undergo a change of several per cent, and the probe is convenient to use in adjusting a klystron or other oscillator for maximum output. They are useful in showing the presence of waves, or whether an ultra-high-frequency generator is actually in oscillation. A probe may readily be built to respond to a power of a few milliwatts.

There are at least three properties of a wave guide that are of practical importance. One is its ability to propagate wave power from one point to another. Under reasonably favorable circumstances, wave guides introduce far less loss than conventional wire lines of either the parallel-wire or coaxial variety. A second property is a necessary consequence of the first and relates to resonance. If a short piece of wave guide is bounded at one end by a metal barrier which may take the form of a tightly fitting but movable piston, and bounded at the other by a fixed plate with a circular iris an inch or two in diameter, the chamber so formed may be made to resonate. Roughly speaking, resonance occurs when the length is an integral number of (guide) half wavelengths. This simple half-wave relation will, however, be modified greatly by the size of the iris chosen. This second property is in effect a reactive property. It makes it possible to use such a cavity as the high-frequency counterpart of the simple tuned circuit consisting of a coil and condenser. In one form it may be given a calibration and be used as a wave meter. In another it may act as a matching transformer between two guides of differing characteristic impedance, while in a third it may either match a generator to a line, thereby presenting to it a more effective load, or it may match a line to a receiver, thereby impressing on it a maximum of available wave power. A more generalized statement would, of course, be that such a chamber can be used to match a source to a load.

The third useful property of a wave guide is its radiating property. One example is the open end of a wave guide. A portion of the wave power arriving at the open end of a pipe proceeds into the free space beyond and is lost to the guide. If we flare the open end into a horn of suitable proportions it enhances radiation and may at the same time launch the same with considerable directivity. If we like, we may regard the horn as a device that tends to match the guide to the outside medium. Southworth and King have published[18] the directional patterns which appear in Fig. 14-13 and serve to illustrate the use of a horn as a radiator at high frequencies.

14-7. Practical Aspects and Applications.[19]- In applications where attenuation is of very great importance, it is desirable to use copper as the guide material. However, in many cases, and especially for short runs, the metal that works easiest mechanically (brass) is probably best. If necessary, silver plating may be added at a moderate cost. It is, of course, preferable to work with standard materials and dimensions. For 3000-mc (λ = 0.1 meter) waves it is convenient to use circular brass pipe of 3" outside diameter with 1/16" wall. For rectangular pipe, either 1-3/4" or 1-1/4" x 2-1/2" (outside measurements), each with 0.081" wall, is suitable. Outside dimensions are specified because they are the ones listed by most American manufacturers.

It has already been pointed out that a short section of

---

18. Southworth, G.C., and A.P. King, Metal Horns as Directive Receivers of Ultra-Short Waves, Proc., I.R.E., 27, No. 2, Feb. (1939).

19. Virtually all the rest of this chapter is an exposition of the properties of hollow wave guides and resonators, presented in the form of a discussion of experiments illustrating these properties. The material follows substantially a note prepared by Dr. G.C. Southworth for the u-h-f techniques course. The authors are glad to acknowledge here their indebtedness to Dr. Southworth for this material.

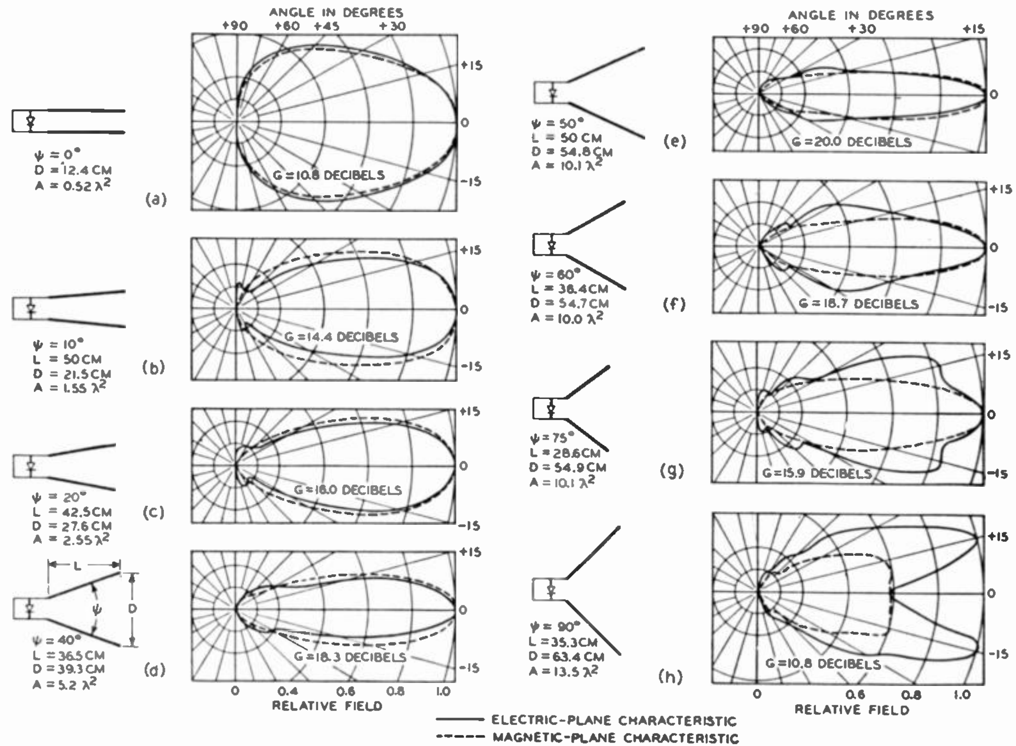


FIG. 14-13. Directivity patterns of metal horns, taken at $\lambda = 15.3$ cm. (Southworth and King, Proc. I.R.E., Feb. 1939.)

wave guide suitably bounded by metallic barriers may be made to resonate and may thereby function in a variety of very useful ways. This is, of course, quite analogous to a similar use sometimes made of Lecher wires or of tuned coaxial lines. Without going into a lengthy discussion of the theory of this process we will proceed with a description of various ways by which its principles may be demonstrated.

Apparatus suitable for this purpose consists of a piece of 3" circular brass pipe perhaps a foot long. In one end is fitted a circular metal piston that may be moved back and forth over a distance of perhaps 8" preferably by a rack and pinion. At the other end is fastened a plate of brass having a central iris perhaps an inch in diameter.

One very simple way to make a test is to bring this resonator within a foot or two of a 10-cm oscillator and mount close to the iris the hand probe already described. Under this circumstance the crystal detector current may be 100 microamperes. If now we slowly move the piston along the guide we obtain a point where the detector current takes a sudden drop. This is one of the resonance points. Moving the piston further along one may obtain a second position in the standing wave pattern. The intervening distance is, of course, a half wave as measured within the guide. The corresponding wavelength in free space may be calculated; see Table 14-I.

By making a slight modification in the above chamber we may improve its accuracy very considerably. To this end we replace the rather crude external indicator by an internal indicator. The latter is a form of the above probe mounted on the outside but having a small pick-up wire that reaches inside the chamber to bring out for indication purposes a very small part of the total available power. In this case we locate the resonance points by a maximum reading rather than by minima. We visualize this indicator as a high-impedance device and accordingly try to locate it near the voltage maximum inside the chamber. This will be approximately a quarter wavelength back of the iris. We must, of course, keep in mind that this wavelength λ_g is the one measured inside of the guide. For the case at hand this quarter-wave distance is about 4 cm.

A probe suitable for this use will extend about 1/4" into the guide. Carefully made readings of wavelength may be expected to agree with calculation to within a few tenths of a per cent. A resonant chamber suitable for the purpose is shown schematically in Fig. 14-14.

The Q factor[20] for the above chamber will depend on the size of the aperture, and the losses incurred in the walls. Their relative importance will be roughly that of the order

20. Defined in Ch. 10, q.v.

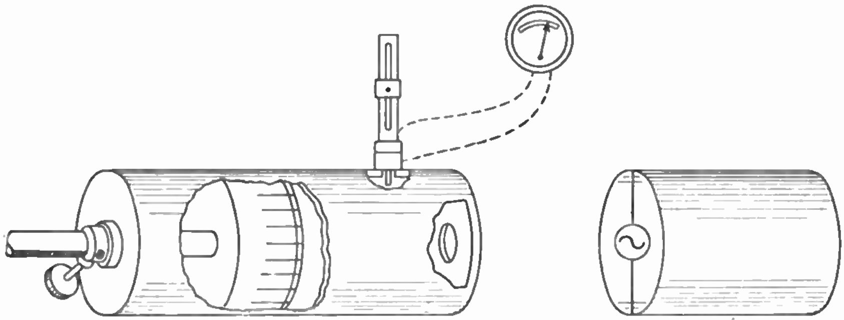


FIG. 14-14. Resonant chamber with internal probe.

presented. The Q of such a chamber may easily be made 1,000. By careful design, it may be extended to 10,000 or even higher. It is obvious, of course, that a very satisfactory resonant chamber may also be built from rectangular pipe.

Some interesting experiments may be built around a chamber of this kind. It is instructive, for example, to plot resonance curves for the cases where the end of the pipe is open, where it is closed by a large iris, where it is closed by a small iris, and where irises of intermediate size are used. One may also place diametrically across the chamber at a voltage-maximum, rods of dielectric, say polystyrene and bakelite, and plot a resonance curve for each. This will show the relative losses produced by these materials and should promote a better appreciation for good dielectrics.

The Traveling Detector. Another particularly useful piece of apparatus has, for the want of a better name, been called a traveling detector. It consists of a probe like that just described mounted on a slider free to move along a longitudinal slot cut in a wave guide. The travel may well be as much as 16" with a length of pipe of 20". There is illustrated as Fig. 14-15 a very simple form of this apparatus that is probably adequate for most experimental purposes. If the slot were so located as to interrupt the flow of currents in the wall of the guide as shown in Fig. 14-16a the result would, of course, be serious but when it is located as shown in Fig. 14-16b it has little effect on the waves within. The latter condition is, of course, that in which the traveling detector is used. The traveling detector affords a means for examining the relative magnitude of any standing waves that may be present in a wave-guide system. Since standing waves mean oppositely directed

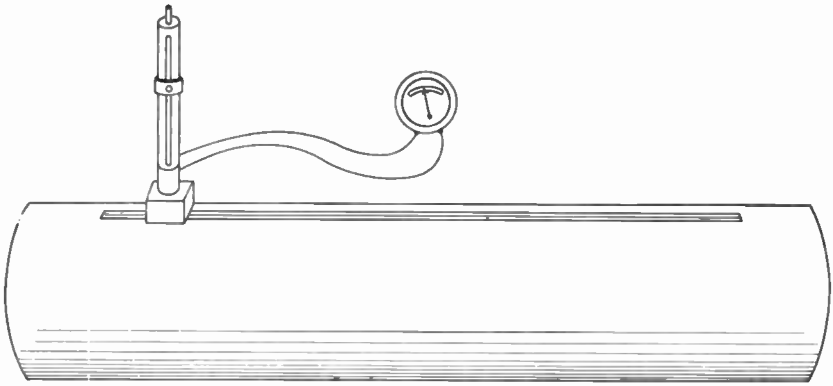
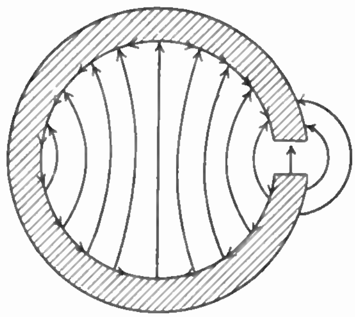


FIG. 14-15. Simple form of travelling detector.

waves and in most cases reflection, this apparatus permits us to test for the presence of electrical discontinuities and, in addition, we may measure the coefficient of reflection. Under certain circumstances the apparatus may also be used to measure passing wave power.

If the detector itself follows a square law or some other known relation, we may derive from its meter readings the relative amplitudes corresponding to the peaks and valleys of the standing wave. Since the first represents the sum of incident and reflected waves and the second their difference, we may derive in arbitrary units their respective amplitudes. Their inverse ratio is the reflection coefficient. If by some means this factor can be made zero it is fair to say that we have terminated the wave-guide line in its characteristic impedance. It is obvious also that if the load that terminates this guide can be made to measure quantitatively the wave power received, we may then calibrate our traveling detector at each of several levels and thereafter be able to tell how much power is passing.

There are several interesting simple experiments that may be built up around the traveling detector. For example, there is shown in Fig. 14-17 a set-up consisting of an oscillator A spaced perhaps a foot from the open end of a wave guide B. The significance of the horn shown dotted will follow shortly. A traveling detector of the kind just described is shown at C. At the remote end D of the traveling detector there may be attached various devices whose reflection effects are to be determined. The first may be a solid disc of metal. Its coefficient of reflection will be substantially unity. The second


(a)

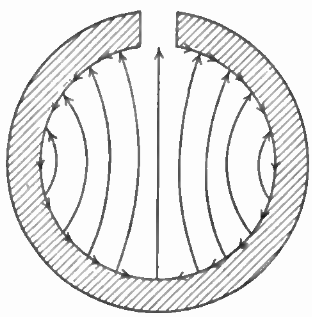

(b)

FIG 14-16 The effect of slots upon transmission of the dominant wave; lines across the guide indicate the electric force, arrows in the inside wall indicate conduction currents.

may be a similar disc with a one-inch iris. The coefficient of reflection may still be high but not unity. These may be followed by progressively larger irises until the open end of the pipe is used. In this case reflection will be small but not zero. Finally a small horn may be attached and the reflection will be found to be substantially zero. The latter will be convincing proof of the efficacy of a horn as a termination. The irises used in this experiment may very well be those used with the resonant cavity above.

With the traveling detector at hand it is possible to test the degree of electrical smoothness of twists and bends that may be placed in a wave-guide line or the smoothness of such circuit elements as matching units or terminating receivers. In fact it now appears that in assembling for the

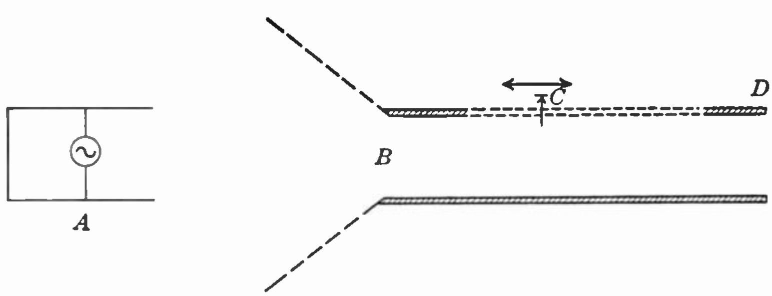


FIG. 14-17. Simple setup for measuring coefficients of reflection.

first time a rather complicated system of microwave components an engineer will need to verify in advance that no serious discontinuity prevails in any particular part of his system.

Matching Units. If we were to place beyond the traveling detector the sharply resonant chamber described earlier we would probably find that there would be no very stable adjustment at which it would terminate the line. If, however, we were to place in the cavity an absorber such as the rod of bakelite used earlier or possibly a substantial amount of slightly moistened cotton we would have little difficulty in choosing an iris diameter and a piston adjustment that would provide a termination. In this instance it would be proper to regard the chamber as a transformer whereby the absorbing cotton (the sink) is matched to the wave guide (the source). In this case it is not difficult to see how reflected waves passing back and forth through the cotton could finally be absorbed.

If we were to repeat this experiment, this time replacing the mass of cotton by a transverse disc of bakelite carrying a thin coating of colloidal carbon, we would also find that the line could be terminated. Here again it is not difficult to see how the waves by their repeated excursions through the disc might also be absorbed. It is not, however, so easy to see how the mass of cotton or the disc could be replaced by a single transverse filament or by a pin point of absorbing material and still be a good termination. Yet traveling detector measurements of the kind just described show that under certain circumstances this can be done.

It is perhaps also easy to see how two end walls, one with an iris, can provide the conditions for a resonant section of wave guide. Again it may be correspondingly difficult to see how these end walls may be replaced by single diametral rods. This also is substantially true. In the latter case, however, it is generally necessary to provide means for tuning.

It is not only possible to replace the end plate containing the iris by a simple tuned diametral rod but we may incorporate with it the absorbing element into which we wish the power to flow. The latter may very well be a crystal detector having a few hundred ohms resistance and be capable of telling us the relative power absorbed. Figure 14-18 shows in schematic form such a terminating receiver. As already implied, standing wave tests have shown that it functions substantially as described. In this case the reflecting end wall is fixed at a point found to be approximately correct by previous experiment. Final adjustments are made by means of trimming screws shown. Tuned rods are, in general, much easier to adjust them irises, and for many laboratory purposes, they are correspondingly more convenient. Figure 14-19 shows a modified form of this receiver in which the crystal detector has been replaced by a

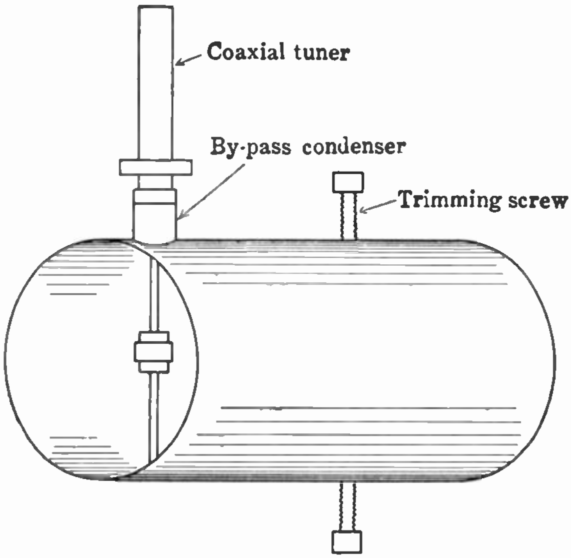


FIG. 14-18. Tuned wave guide receiver employing a rectifying crystal.

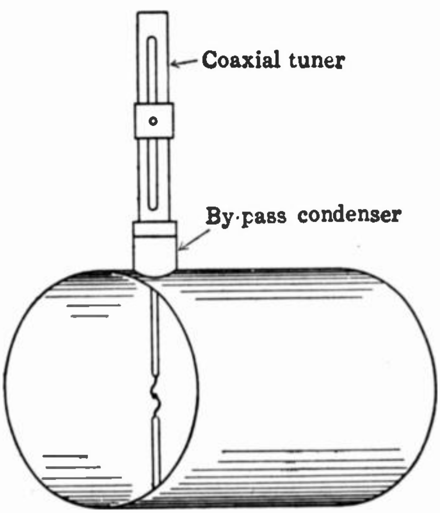


FIG. 14-19. Section of wave guide with a tuned uranium oxide unit.

small crystal of uranium oxide. This kind of an absorber is even smaller than a silicon detector. Yet it seems to function equally well as an absorber.

Uranium oxide is not ordinarily used as a rectifier or as a demodulator. However, it may provide a very convenient method of measuring power. This material has a very high negative temperature coefficient of resistance. If we place it in a microwave circuit as shown, it will become heated by as little as 10 milliwatts of power, and show a corresponding change of resistance when connected externally to an ohmmeter or a Wheatstone bridge. If next we replace the high-frequency power by the necessary amount of measured low-frequency power to produce the same resistance change, we then have a measure of power.

Matching a Wave Guide to a Coaxial Line. In the matching units just shown, a crystal is the load. It is a matter of experience that such a load may be replaced by a coazial conductor that leads away the power. Figure 14-20 shows a practical arrangement for bringing this about. Tests with a traveling

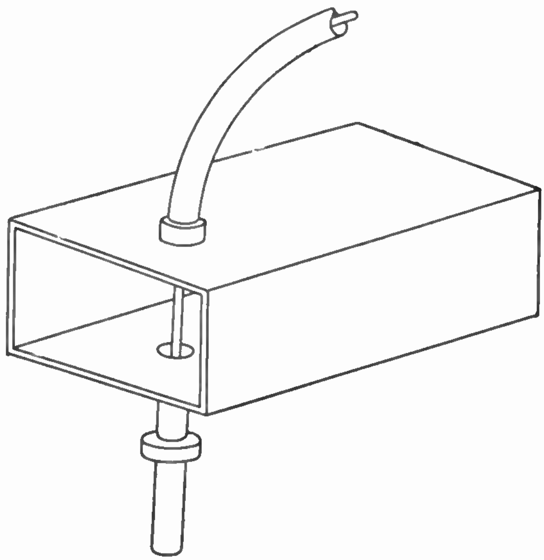


FIG. 14-20. Matching unit between wave guide and coaxial cable.

detector placed in the wave-guide line will again verify that the arrangement is electrically smooth. Figure 14-21a and b show other arrangements for accomplishing the same result. The first is extremely simple and is entirely adequate for many purposes. The second involves a complexity that seldom can be justified.

The fact that coaxial lines may be made flexible and of relatively small diameter, as well as their shielding property and other characteristics, have made their use rather general in microwave work. No doubt they will continue to be used, particularly where attenuation losses are unimportant. It now appears that ultimately microwave installations will involve a mixture of the two kinds of connections. For purposes of comparison and for information for use in deciding between the two, the following data may be used. A rectangular wave guide of copper specified above as 1-3/4" x 3" O.D. has an attenuation

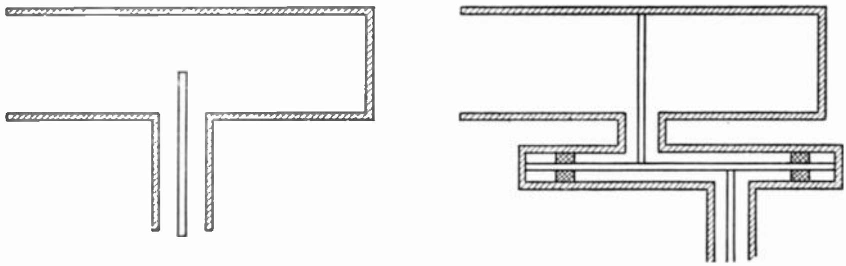


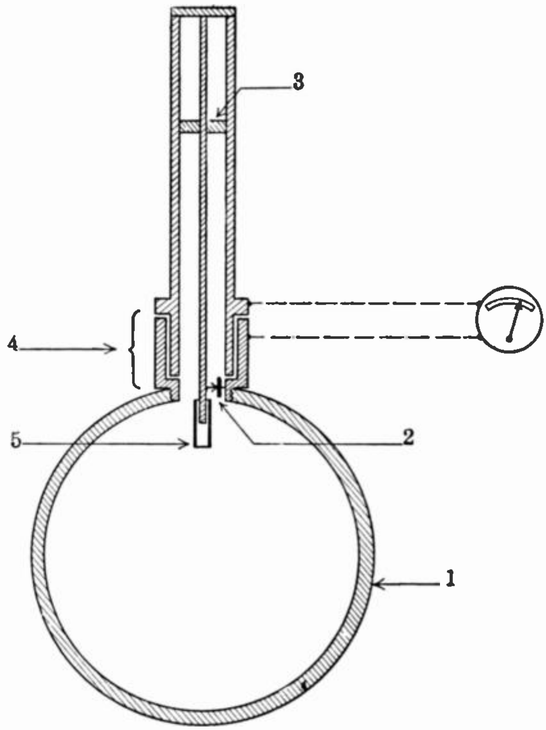
FIG. 14-21. Matching units between wave guide and coaxial cable.

at λ = 0.1 meter of 0.0175 decibel per meter. A form of flexible coaxial line incorporating a continuous loading of rubber and suitable for plug-and-jack connections has an attenuation at the same frequency of 5 db per meter. Another form of cozaial somewhat less rugged has an attenuation of 0.7 db per meter, while in a third rather special form this has been reduced to 0.4 db per meter. By sacrificing flexibility, this latter figure can probably be given a further slight improvement. However, the principal advantages of such a line will then be lost.

Detector and Coaxial Tuner Components. It will, of course, be a great convenience if the coaxial tuner and detector arrangements described above can be carried into effect with a minimum of apparatus. This may be possible if we adopt some standardization in designing the various components. A few items that have been found to work well are described below. Before going into their details, however, it may be well to review again the various ways in which such components are used.

Figure 14-22 shows, in schematic form, the tuner and probe detector arrangement needed for resonant-chamber or traveling-detector measurements. It will be noted that the crystal is placed across the tuner and a by-pass condenser is provided. The latter may be made a part of the detector mounting. Rectified currents or other demodulation components may be taken off by means of the meter connections shown dotted.

This form of detector is referred to as the shunt type. Its mounting is shown in greater detail in Fig. 14-23. Figure 14-24 shows the corresponding arrangement needed when the detector and its associated tuner is used as a termination. This form of detector is for obvious reasons called a series type. Its mounting is relatively simple but an external by-pass

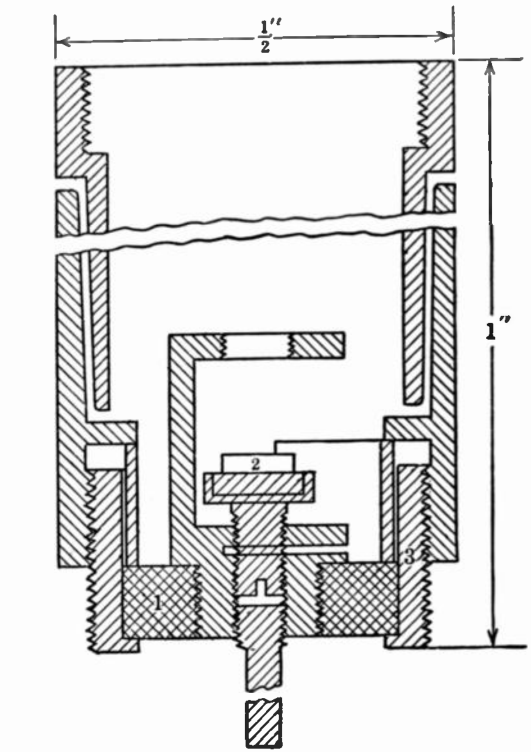


1 – wave guide
3 – tuning piston
2 – crystal detector
4 – by-pass condenser
5 – small tubing over rod

FIG. 14-22. Arrangement of coaxial and detector components in a tuned probe.

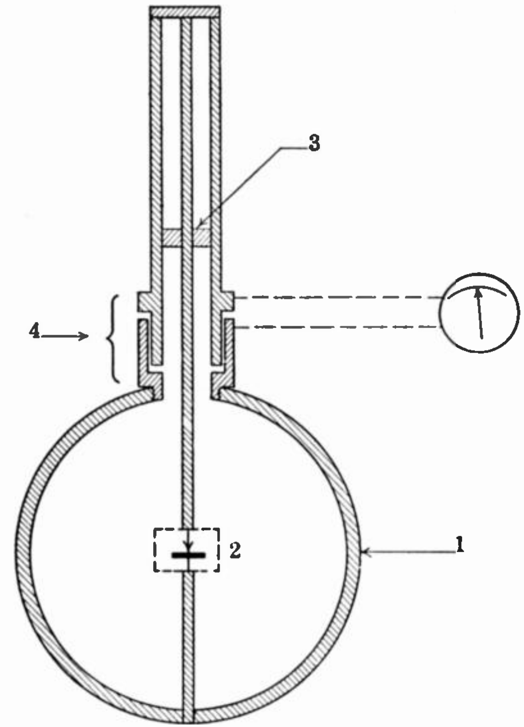
condenser is needed. Details of the crystal mounting are shown in Fig. 14-25. The by-pass condenser follows the general lines of that built into the shunt type. See Fig. 14-23. It is convenient to grind the crystal used in these devices into a small cylinder and fasten the same into a mounting using a set screw or preferably solder as shown in the above sketches.

There are many materials that may be used as detectors. For measurement work, it is a very great convenience if the rectified current is proportional to the square of the applied voltage. For cases where sensitivity is more important than is convenience, as, for example, in the reception of weak signals, a material whose characteristic exponent is higher than two is preferable. Crystals of iron pyrites usually follow the square



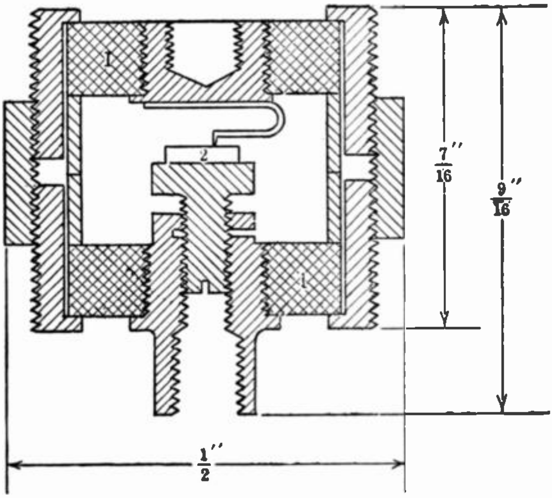
1 – ceramic insulator   2 – crystal

FIG. 14-23. Detector Assembly— Shunt Type.



1 – wave guide   3 – tuning piston
2 – crystal detector   4 – by-pass condenser

FIG. 14-24. Arrangement of coaxial and detector components in a terminating receiver.



1 – ceramic insulators 2 – crystal

FIG. 14-25. Detector Assembly – Series Type.

law rather closely. If this material is available it will probably be very satisfactory for measurement work.

Silicon has for many years also been used as a detector. Its rectifying characteristics apparently depend largely on conditions of manufacture. These may be controlled so as to bring about a variety of results. Two types of synthetic silicon detectors have been evolved for use. In one type the square law is followed over a considerable range. In the other a higher exponent is obtained.

In determining the characteristic exponent of a detector it is often convenient to set it up in a terminating receiver such as shown in Fig. 14-18. This is then coupled to a wave source and oriented until the output is a maximum. Under this circumstance the diametral leads to the detector should lie along the lines of electric force. The cylindrical receiver is then rotated about its major axis as data on rectified current and angle are taken. It is assumed that applied voltage is proportional to the cosine of the angle. The results may conveniently be plotted on semi-log paper.

Figure 14-26 shows a simple form of coaxial tuner that is sufficient for many laboratory purposes. It may be made of brass, but contact troubles may develop later when the brass tarnishes. The latter may be more or less obviated if the moving parts are silver plated. The product will be still better if it is made of coin silver. In the design shown, the shorting piston is adjusted by means of an external knurled ring connected to the piston through a longitudinal slot. In a modified design, capable of much more accurate settings, the piston is operated by means of a lead screw.

Mode Filters. It is difficult or impossible to set up initially in a hollow guide a single pure mode of oscillation. However, after producing the excitation in the way best suited to the mode desired, a filtering out of the other modes may be accomplished in two ways.

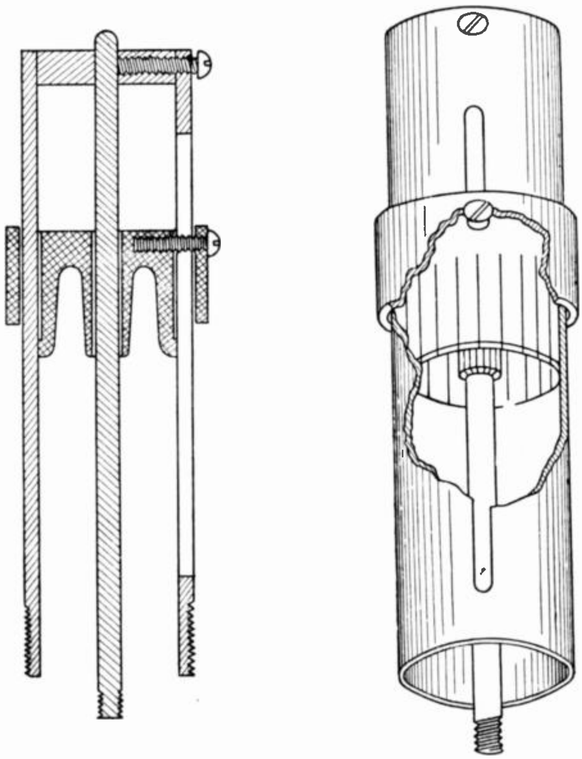


FIG. 14-26. Simple form of coaxial tuner.

First, the size of the tube may be selected so that the frequency used is above the cutoff frequency for the first mode, but slightly below the cutoff frequency for the second or next higher mode. Then if any higher modes exist near the exciting antenna, they will be confined to its immediate neighborhood because of the cutoff characteristic. Similarly, undesired modes will be suppressed if originated at bends or other discontinuities.

If there is present in a guide an unwanted wave having an electric field $\underline{E}$ so oriented that it has a non-zero value along certain lines in the tube where the desired wave has zero value, a number of thin metal wires may be disposed in the tube, coinciding with these lines, and they will eliminate or reduce materially the unwanted mode. For example, if the $TE_{0,1}$ mode is to be used in a round pipe, but some energy in the $TM_{0,1}$ mode is found to be present, there may be inserted in the pipe a "mode filter" consisting of a set of radial wires in a transverse plane, affixed to the inside of the tube and pointing toward its center, but not touching one another there. It is seen from eq. (14-69) that $E_\rho = 0$ for the $TE_{0,1}$ mode, and hence thin radial wires will have but little effect on the passage of the wave. On the other hand, for the $TM_{0,1}$ wave the E_ρ component, eq. (14-63a), is not zero, and this component of the field will be effectively short-circuited by the thin radial wires.

REFERENCES ON WAVE GUIDES

"On the Passage of Electric Waves through Tubes, or the Vibrations of Dielectric Cylinders," Lord Rayleigh, Phil. Mag. 43, 125-32 (1897). Also in "Scientific Papers," 4, 276-82.

"Transmission of Electromagnetic Waves in Hollow Tubes of Metal," W. L. Barrow, Proc., I.R.E., 24, 1298-1328 (1936).

"Hyper-Frequency Wave Guides--General Considerations and Experimental Results," G. C. Southworth, Bell Sys. Tech. Jour., 15, 284-309 (1936).

"Hyper-Frequency Wave Guides--Mathematical Theory," J. R. Carson, S. P. Mead, and S. A. Schelkunoff, Bell Sys. Tech. Jour., 15, 310-333 (1936).

"Electromagnetic Waves in Hollow Metal Tubes of Rectangular Cross Section," L. J. Chu and W. L. Barrow, Proc., I.R.E., 26, 1520-55 (1938).

Chapter 15

LABORATORY MANUAL

The group of experiments outlined in brief form below is intended to illustrate some of the more important and basic principles and techniques which have been discussed in this text. It is assumed that the student has already been introduced to a small amount of communications laboratory work and equipment. Succinctly, he should have already performed experiments covering substantially the following topics or their equivalent:

(a) Series and parallel resonant circuits.

(b) Impedance bridge measurements and general technique in the use of bridges and similar equipment at audio frequencies.

(c) Static characteristics of vacuum tubes.

(d) Dynamic characteristics of vacuum tubes.

(e) Construction and test of a power supply (Experiment 1 below is a more detailed study than assumed here).

(f) Technique of use of cathode-ray tubes (CRT's) and cathode-ray oscillographs (CRO's).

(g) Audio-frequency amplifiers and oscillators.

With this background it is considered feasible to proceed with the experiments which are described below. Their titles are

1. Power Supplies
2. Sweep Circuits and Electronic Switching
3. Electrostatic and Magnetic Deflection of Cathode-ray Tubes.
4. Negative Feed-back Amplifier
5. Square-wave Response of Circuits
6. Detectors--Crystals, Diodes, Pentodes
7. Video Amplifiers
8. Superheterodyne Receiver, Using the RCA Demonstrator
9. Negative-grid Oscillator; Lecher Wires

---

1. Most of the experiments outlined later follow the conference outline and are taken specifically from the group of experiments performed in the Moore School of Electrical Engineering of the University of Pennsylvania. The laboratory was in charge of Dr. Carl C. Chambers, who was assisted by Messrs. Iredell Eachus, J.P. Eckert, and Bradford Sheppard. Professor Reich supplied directions for two experimente. It is planned in a later edition to make the laboratory manual a summary of the experiences of all the authors, and of such other persons as contribute criticisms.

10. Magnetrons and Klystrons
11. Impedance Matching with Lumped Constants
12. Antenna Feeding and Matching with Stubs
13. Antenna Arrays
14. Hollow Wave Guides--Part I
15. Hollow Wave Guides--Part II
16. Horns and Parabolas.

Throughout his laboratory work the student may well endeavor to absorb numerous practical facts and techniques not specifically mentioned in the experiments. Listed below are many such points.

CIRCUIT ELEMENTS, PARTS AND PRACTICAL CONSTRUCTION

Resistors - residual L and C; types of construction (wire wound, composition, metallized); power dissipation; temperature limitations; physical size and appearance of modern types.

Condensers - residual L and R; types of construction (mica, paper, electrolytic, air, oil); power dissipation; temperature, pressure and humidity effects; voltage breakdown; realizable power factor; physical size and appearance of modern types.

Inductors - residual C and R; skin and proximity effects; types of construction (air and iron cores, solenoid, toroid and other winding forms); power dissipation and current capacity; temperature and humidity; voltage breakdown; realizable Q; physical size and appearance of modern types.

Batteries - dry cells (internal resistance, deterioration under extremes of use and shelf life, temperature effects); storage cells (internal resistance, charge and discharge cycles, specific gravity); physical size, weight and appearance of modern types.

Wire- insulation (cotton, silk, rubber, enamel, glass, asbestos); current capacity and voltage breakdown; pushback and spaghetti; resistance for d-c. and a-c. through skin effect range; shielded wires and cables; cable connectors.

Solder - composition (soft, hard, eutectic, silver); fluxes (acid, rosin, non-corrosive pastes); flux solvents;

restrictions as to metals; soldering tools; technique of soldering.

Chassis Assembly - materials of chassis (copper, aluminum, steel, chromium plate); shields and shield cans; shielded wire; grounding; terminal strips and binding posts; insulators; chassis layout and wiring considerations; self-tapping screws.

Use of Tools - drills (hand and power) high and low speed, lubricants; reamers; punches; spot welders; metal saws, hand and power; shears; thread cutting and thread cutting tools; simple lathe and miller work.

CONVENTIONAL DEVICES AND TECHNIQUES

Ammeters - d-c types; power-frequency types; rectifier types; thermocouple types.

Voltmeters - above types of ammeters modified for use as voltmeters; electrostatic types.

Power Meters - power-frequency wattmeters; rectifier type output meters, thermal types.

Frequency Measurement Techniques - heterodyne method; Lissajou's method; calibrated resonant circuit methods; electronic frequency meters.

Bridges - wheatstone d-c types; impedance (1000-cycle) bridge; capacitance bridge; grounding, shielding, bridge amplifiers, and balancing technique.

Test Oscillators - simple types; standard signal generators; alignment sweep oscillators.

Tube Testers - vacuum tube bridge; emission and transconductance testers.

Cathode-ray Oscilloscopes and Their Techniques.

Square-wave Generators and Testing Technique.

Q Meter.

U-H-F DEVICES AND TECHNIQUES

Ammeters - thermocouple types; bolometer types; crystal detector types.

Voltmeters - thermocouple types; diode peak-reading types; electrostatic types; crystal detectors.

Power Meters - thermocouples; bolometers; calorimetric methods; photometric method.

Frequency and Wavelength Measurements - heterodyne methods; resonant circuits and absorption methods; coaxial-line and cavity type wave-meters.

Impedance Measurements - method employing standing waves on transmission lines or wave guides; modifications adapted to different wavelength ranges; applications to several types of problems; transmission-line bridge.

Q Measurements - method employing response curve of resonant circuit.

Radiation Patterns - crystal detector or bolometer; rotation of antenna structure; reciprocity and transmitted or received pattern; concepts of beam angle, lobe size and power gain.

Spectrum Analysis - method of measuring energy distribution of oscillator using high-Q resonator as wavemeter.

15-1. Power Supplies.- (a) Using an unregulated full-wave vacuum-tube type rectifier circuit (to be designed and constructed by the student from available parts), determine curves of ripple voltage vs load current and d-c output voltage vs load current (resistance load).

(b) Holding the resistance load fixed, determine the curve of d-c output voltage vs a-c line input voltage. Vary the latter from 70 to 130 volts by use of an autotransformer.

(c) (CAUTION: In this part of the experiment place a large (order of 10 μf) capacitor[2] in series with the power supply and bridge, in order to prevent the large d-c voltage of the power supply from damaging the bridge. Do not neglect the effect of this capacitor on the results.) Using the General Radio Impedance Bridge Type 650A or an equivalent, measure the impedance looking back into the power supply, for numerous types of load and numerous load currents.[3]

2. Preferably variable, in order that a value can be chosen such that the capacitance will not completely neutralize the inductive reactance of the power supply. A Cornell-Dubilier variable capacitor works well.

3. Since the load is presumably known, the actual output impedance of the power supply can be found from the measurement.

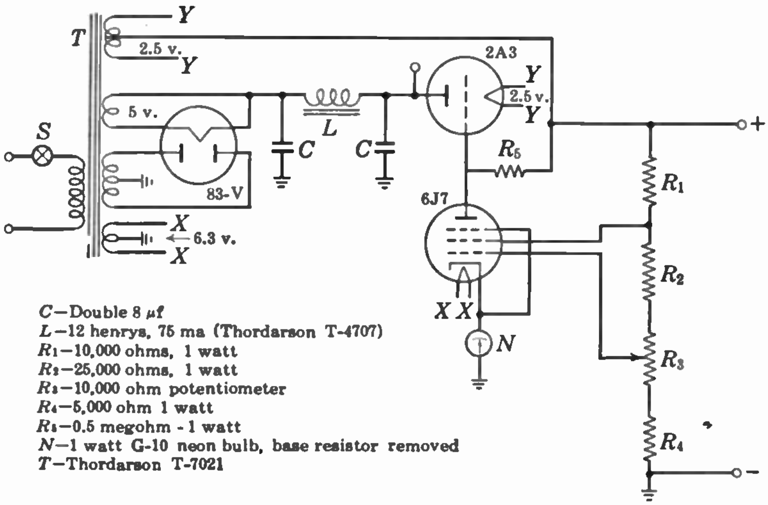


FIG. 15-1. Voltage-regulated power supply. (If *N* does not glow, regulator is not functioning).

(d) Repeat parts (a) to (c) inclusive using a regulated power supply of the type shown in Fig. 15-1 or an equivalent. In measuring the ripple voltage it will probably be necessary to insert an amplifier between the output terminals of the power supply and the oscilloscope used for making the measurement. (Note. If the student has not carried out an experiment comparing various types of rectifier filters, as discussed in Ch. 2, this should be done before the experiment described above.)

15-2. Sweep Circuits and Electronic Switching.- Plot all wave forms obtained in this experiment.

(a) Connect the necessary external d-c and a-c voltages, meter and variable (commutating) condenser box to the thyratron trigger circuit or inverter. (See Fig. 15-2.) Remove the 2A4G thyratron which has its plate connected to one side of the switch S_2. The remaining thyratron is fired by closing the switch marked S_1 momentarily. It can be extinguished by closing switch S_2 for a short time. Explain these actions. Plot a curve of minimum value of commutating condenser for extinction against load current. Load current is varied by the step load resistor.

(b) Replace the 2A4G thyratron. Fire one thyratron, then the other, by alternately closing switch 1 and switch 3. Explain. Why will the curve plotted in part (a) not apply here? Plot the corresponding curve for this case.

(c) Using jumpers, connect the a-c grid excitation

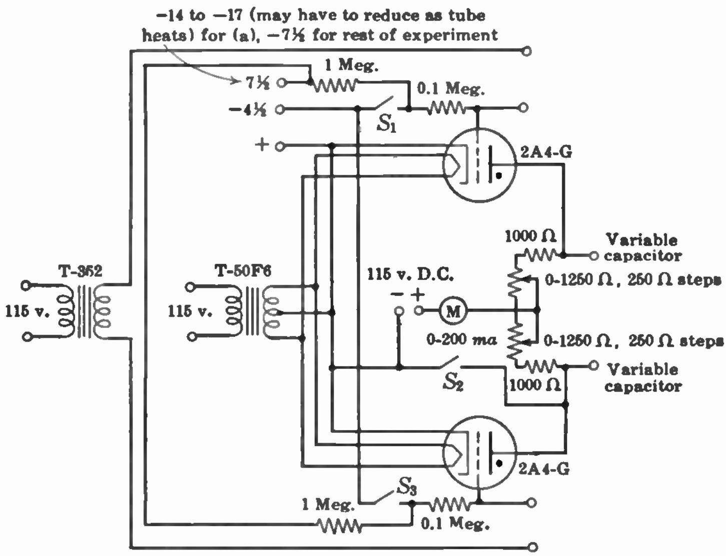


FIG. 15-2. Thyratron trigger circuit.

transformer to the grids of the thyratrons. Turn on the grid excitation. Positive pulses are applied to each grid circuit on alternate peaks of the a-c voltage wave. Examine the wave forms across the plate resistors with a cathode-ray oscillograph. The outputs of the two tubes may be simultaneously observed in the proper phase relation by using the electronic switch. The output of a single tube relative to its grid voltage may also be simultaneously observed in the proper phase relation by use of the electronic switch. (See Figs. 5-18 and 5-19.) Plot the wave forms observed above in the proper phase relation.

(d) Attach a variable capacitor to a relaxation oscillator (Fig. 15-3 shows a possible circuit). Connect a cathode ray oscillograph to the circuit using the electronic switch to obtain simultaneous wave forms of the grid voltage and plate voltage with the circuit oscillating. Note the effect on amplitude, frequency, and linearity of different grid voltages, plate voltages, charging resistances and capacitances of the charging condenser. Repeat the above with the electronic switch connected to obtain simultaneous wave forms of plate voltage and plate current.

(e) With the relaxation oscillator adjusted to give a linear sawtooth wave apply the output to the abscissa amplifier of a cathode-ray oscillograph. Apply a sine-wave voltage to the ordinate amplifier. Apply just enough voltage from the above sine-wave supply to the transformer in the grid circuit of the 885 tube to hold the pattern stationary. Increase this voltage by a large factor. Discuss the results using data obtained with the electronic switch and cathode-ray oscillograph.

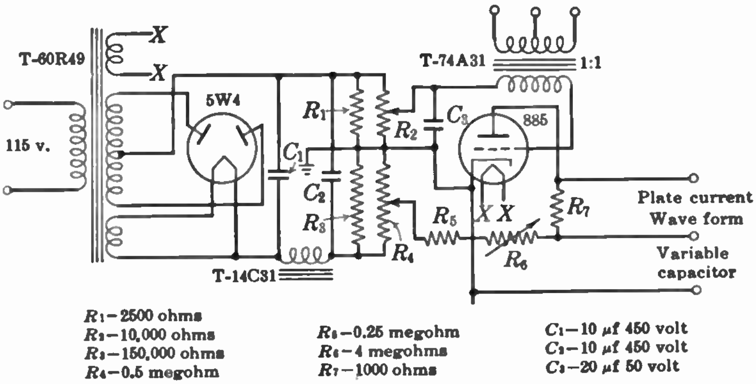


FIG. 15-3. Relaxation oscillator.

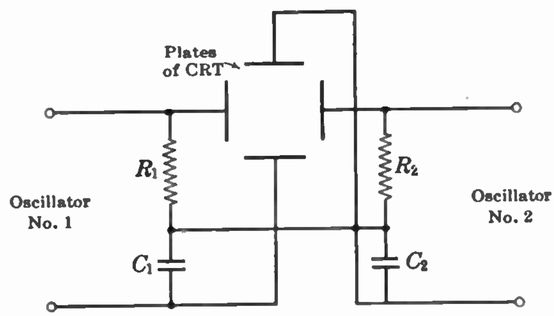


FIG. 15-4. Circular sweep circuit. (Compare Fig. 5-6a).

15-3. Electrostatic and Magnetic Deflection of Cathode-Ray Tubes.- Consult Fig. 5-17 if a Type 208 Dumont tube is to be used, or a corresponding figure for the tube available. It should be possible to make connections to all four deflecting plates.

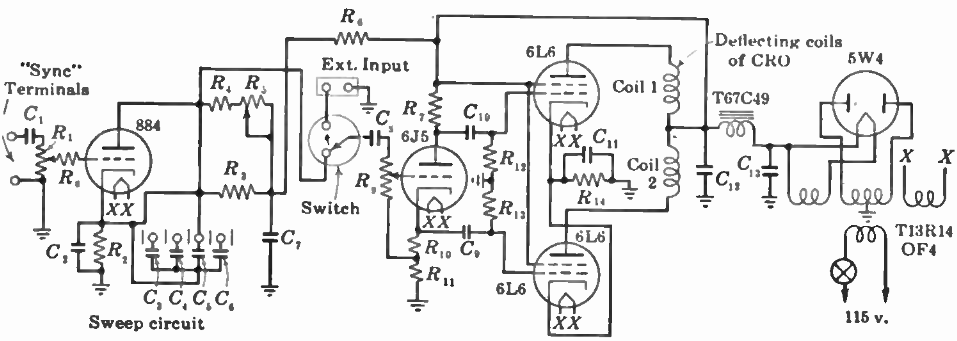
(a) Determine the curve of deflection vs deflecting plate voltage for the abscissa plates and for the ordinate plates of a cathode-ray tube.

(b) Determine the curve of deflection vs deflection coil current (magnetic deflection). (Deflection coils should be placed around tube neck.)

(c) Apply voltages from two oscillators to the plates of a cathode-ray tube using the circuit of Fig. 15-4 (compare Fig. 5-7). Adjust R_1, C_1, R_2, C_2 so that a circular pattern is obtained for each oscillator separately. Observe the pattern obtained when both voltages are applied. Explain, and contrast with Fig. 5-6b.

(d) Calibrate one oscillator against another, using the CRO.

(e) Adjust the magnetic deflection circuit (Fig. 15-5) or its equivalent for a linear sweep. Sketch the wave forms in the various parts of the circuit relative to one another using the electronic switch. Synchronize the sweep circuit to an external signal by applying a signal to the "sync" terminals and making necessary adjustments.



C_1 – 0.1 μf., 400 v.
C_2 – 25 μf., 25 v.
C_3 – 0.2 μf., 400 v.
C_4 – 0.04 μf., 400 v.
C_5 – 0.1 μf., 400 v.
C_6 – 0.0025 μf., Mica
C_7 – 10 μf., 450 v.
C_8 – 0.1 μf., 400 v.
C_9 – 0.1 μf., 400 v.
C_{10} – 0.1 μf., 400 v.
C_{11} – 50 μf., 50 v.
C_{12} – 15 μf., 450 v.
C_{13} – 10 μf., 450 v.

R_1 – 100,000 ohms, pot.
R_2 – 1000 ohms, ½ watt
R_3 – 50,000 ohms, 2 watt
R_4 – 750,000 ohms, ½ watt
R_5 – 4 Megohms pot.
R_6 – 25,000 ohms, ½ watt
R_7 – 10,000 ohms, ½ watt
R_8 – 10,000 ohms, ½ watt
R_9 – 1 Megohm pot.
R_{10} – 1250 ohms, ½ watt
R_{11} – 10,000 ohms, ½ watt
R_{12} – 0.5 Megohm, ½ watt
R_{13} – 0.5 Megohm, ½ watt
R_{14} – 250 ohms, 10 watt

FIG. 15-5. Magnetic sweep circuit and deflection amplifier.

(f) Throw the switch which converts the magnetic sweep circuit into a magnetic deflection amplifier. Using an oscillator and a vacuum-tube voltmeter make a frequency run of this deflection system. Why have critical distance tubes been used in its output? Plot the data and account for the shape of the curve.

15-4. Negative Feedback Amplifier.- Consult Fig. 15-6; also Ch. 3. Adjust feedback for all cases so that oscillations do not occur at any frequency used in tests. Plot the response curves obtained in decibels gain vs frequency on semi-log paper (4-cycle will be needed). Use an oscillograph in the experiment to make sure that the amplifier is not overloaded in any of the tests. An attenuator box between the oscillator and the amplifier will be necessary to prevent operating the oscillator at such a low level that it would introduce a high relative hum level.

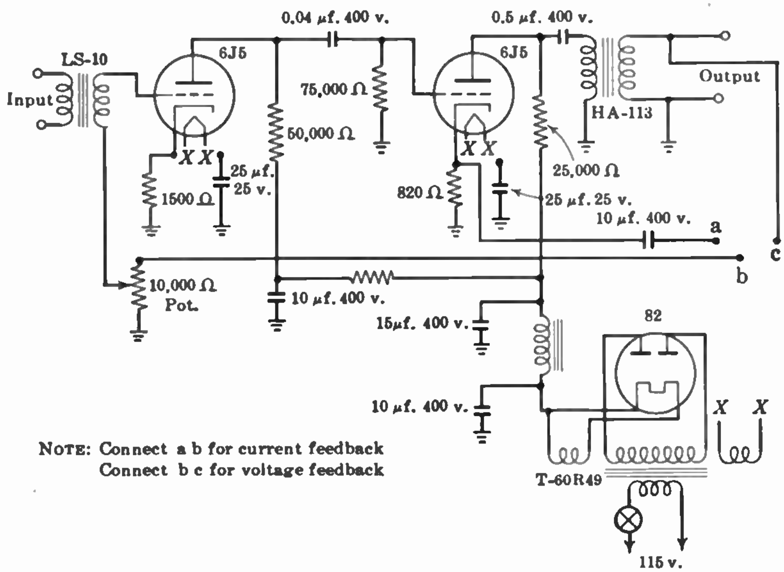


FIG. 15-6. Negative feedback amplifier.

(a) Keeping the input voltage constant in magnitude, measure the output voltage as a function of frequency with the external feedback circuits inactive, for various conditions of cathode by-passing. Measure the output impedance under the above conditions by means of an impedance bridge.

(b) Repeat (a) with voltage controlled feedback for both cathodes by-passed.

(c) Repeat (a) with current controlled feedback for the first cathode by-passed.

(d) By measurement of the resistances in the feedback circuit determine the percentage feedback used in (b) and (c).

(e) Measure the phase shift in the amplifier for all the conditions of operation used above. Plot the phase shift vs frequency characteristic on semi-log paper.

(f) Find the type of feedback, condition of cathode biasing, and setting of feedback control to give the strongest oscillation. (The circuit connections for this are unique.) Do this experimentally, then plot on polar coordinate paper the phase shift vs $A\beta$ for the different frequencies. $A\beta$ is the product of the voltage gain of the amplifier multiplied by the fraction of the output voltage feedback.

(g) Reduce the feedback to that used in (b) and show that the gain for a particular frequency is:

$$\text{Voltage Gain} = \frac{A}{1 + A\beta}$$

where A is the voltage gain without feedback (see Ch. 2). Using the amplifier as in (b) plot the polar curve as in (f).

15-5. Square-Wave Response of Circuits.- The general topic of the square-wave response of circuits is touched on in Ch. 1. It is intended that a square-wave generator such as the Hewlett-Packard Model 210 or equivalent shall be used. The student should obtain the circuit diagram and operating data of the square-wave generator as supplied by the manufacturer and should familiarize himself with its operation and the method by which it obtains square waves. Determine its output impedance and do not neglect the importance of this in using it in the following tests.

(a) Connect the output of the square-wave generator directly to the vertical deflecting plates of a cathode-ray oscillograph. Sketch the wave form at a number of frequencies within the range of the square-wave generator. Determine the range of square-wave frequencies for which the vertical amplifier of the oscillograph introduces noticeable distortion. Using a sine-wave source, determine the "half-power" frequencies (see Ch. 1) of the vertical amplifier in the oscillograph for one or more settings of the gain control. Explain the reason for the effects noted.

(b) Determine the range of square-wave frequencies within which the electronic switch operates satisfactorily. Study its circuit and discuss reasons for its limitations. (See Figs. 5-18 and 5-19.)

(c) Impress a square wave on a circuit comprising a resistor and a capacitor in series. Using the oscillograph (and electronic switch) compare the square-wave input with the voltage across the resistor and across the condenser for several values of each. Let the capacitive reactance at the square-wave frequency (sine wave of same frequency) equal approximately 0.1, 1.0 and 10.0 times the resistance. Sketch the waves obtained.

(d) Using a suitable inductance and resistance make observations similar to those in (c). Sketch the waves obtained.

(e) Observe the square-wave response of several interstage audio transformers. Determine the effect of various source and load resistances. Note the effect of putting capacitance across the primary and secondary. Note also the effect of series condensers in the primary and secondary circuit. Use also an inductive load. Use at least one poor and one good transformer in making the tests. Plot the waves obtained relative to the input wave using the electronic switch.

(f) Observe the square-wave response of a suitably terminated low-pass filter. Plot the waves at various stages in the filter relative to the input using the electronic switch. Repeat the above for a high-pass filter and a band-pass filter. Note figures of Ch. 1.

15-6. Detectors--Crystals, Diodes, Pentodes.- Detectors are discussed in Ch. 7. In (a), (b), and (c) below, a wide range of currents should be used (1μa to 25 ma).

(a) Determine the current vs voltage characteristic of (1) a silicon and (2) an iron pyrites crystal detector, in both directions. Plot the resistance vs voltage characteristics.

(b) Plot the d-c current vs applied a-c voltage using a low impedance audio source, for both the detectors used in (a). (A 60-cycle power line may be a good low-impedance source.) See Fig. 15-7a.

(c) Plot the d-c voltage (across crystal) vs a-c current through the circuit using an audio-frequency source. See Fig. 15-7b.

(d) Determine the rectification characteristic (Ch. 7) of a diode. See circuit in Ch. 7.

(e) Repeat (d) using the pentode detector in the circuit of Fig. 15-7c.

(f) Repeat (d) using an infinite impedance detector (use a high-mu triode) as indicated in Fig. 15-7d.

15-7. Video Amplifiers.- Video amplifiers are treated in Ch. 2. A wide-range oscillator or signal generator, a good vacuum-tube voltmeter, and a suitable attenuator are needed.

(a) Determine the gain--frequency characteristic of an uncompensated two-stage amplifier of the type represented in Fig. 15-8, or a similar one.

(b) With the aid of a cathode-ray oscillograph determine the maximum output voltage obtainable without severe distortion.

(c) Modify the amplifier used in (a) by placing a compensating inductance in series with the load resistor R_1. (First calculate the value of inductance required for good compensation.) Determine the gain-frequency characteristic of the compensated amplifier. Observe the response of the amplifier. Observe the response of the amplifier with square-wave signals of various frequencies.

(d) Repeat the observations of (c) using larger and smaller values of compensating inductance.

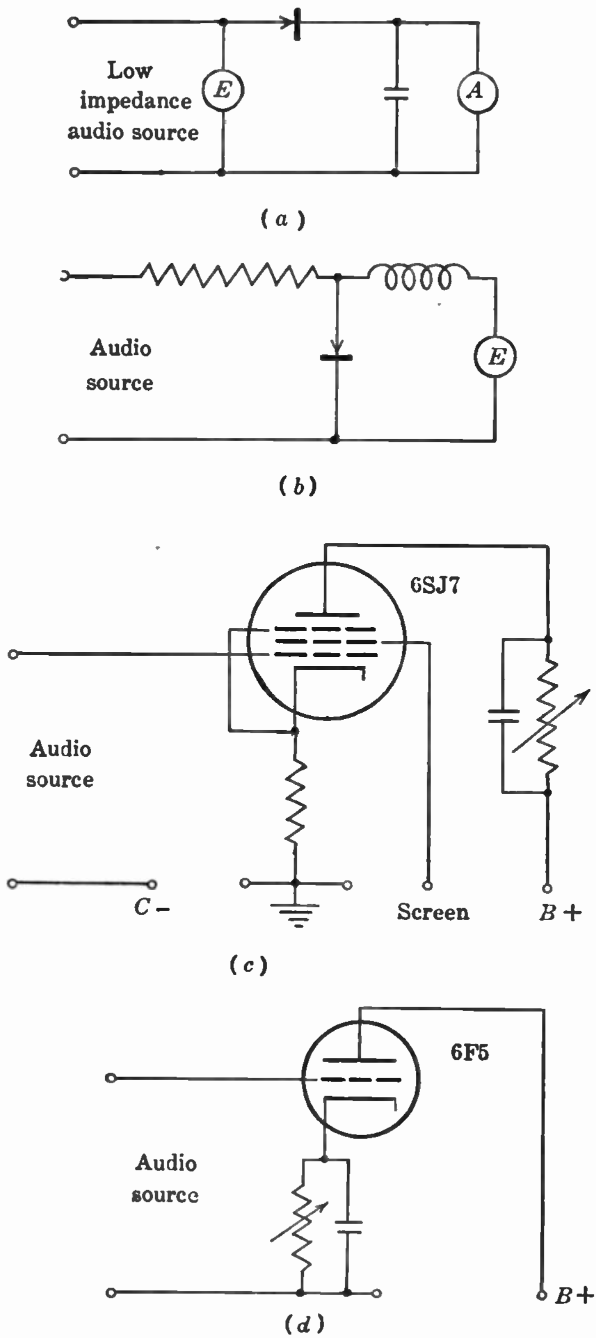


FIG. 15-7. Detector experiment circuits.

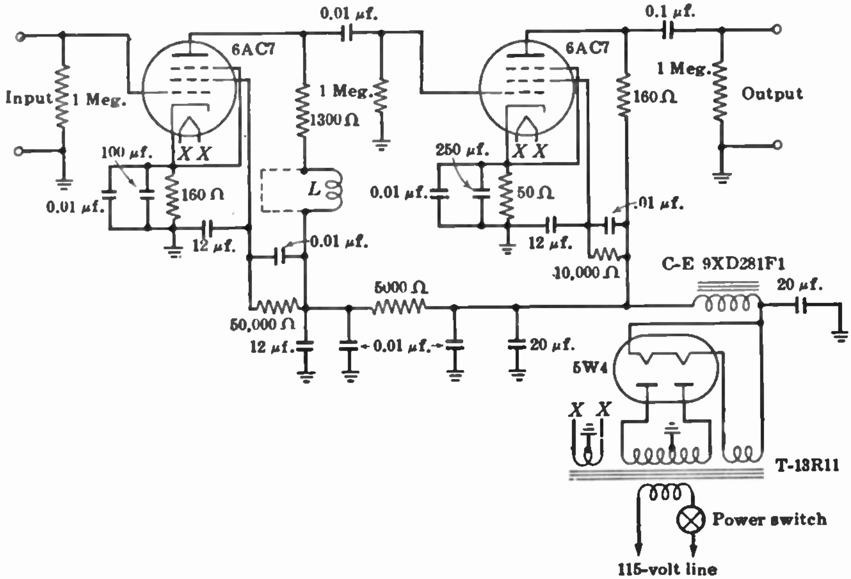


FIG. 15-8. Video amplifier (L = compensating inductance).

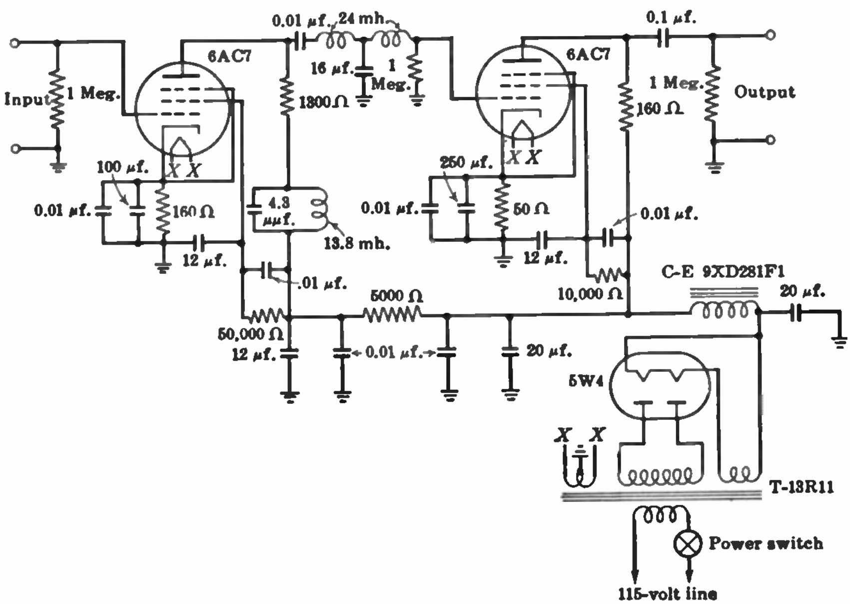


FIG. 15-9. Filter-coupled video amplifier.

(e) Measure the gain-frequency characteristic of a filter-coupled amplifier of the type represented in Fig. 15-9.

15-8. The Superheterodyne Receiver.- The superheterodyne receiver is discussed in Ch. 8 and component parts in various other places of the text. One of the RCA Dynamic Demonstrators may be used (a diagram of Demonstrator II is given in Ch. 8).

(a) Superheterodyne Receiver Alignment. The receiver is to be restored to proper condition after having been made defective beforehand. Use RCA Rider Signalyst and RCA Rider Chanalyst, regular radio service instruments. Give short description of the faults found, method by which they were located, and manner of correcting them.

(b) Sensitivity Measurements (Fixed Bias). With 0.6 volt as reference voltage across the second detector output load, measure the sensitivity of the receiver as a function of the r-f input frequency. Expressing the sensitivity in decibels below 1 volt, plot the sensitivity against frequency.

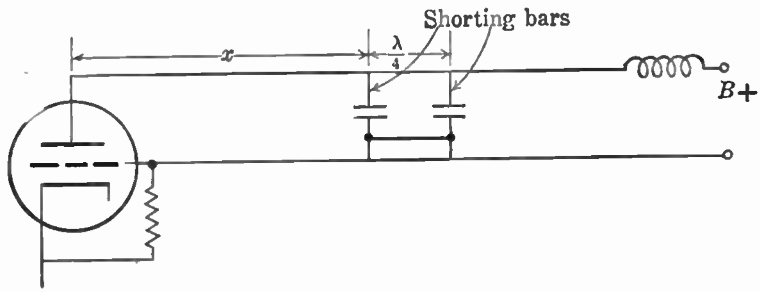
(c) Selectivity Measurements (Fixed Bias). With the receiver tuned to 1400 kc take readings for a selectivity curve, using 0.6 volt across the second detector load as reference. Make measurements at multiples of 0.5 kc off resonance. Plot the selectivity curve by expressing the ratio between microvolts at point off resonance to microvolts at resonance as a function of frequency (or kc off resonance). (As an extension, (c) may be repeated for other frequencies over the tuning range.)

(d) I-f Amplification (Fixed Bias). With the receiver tuned to 1000 kc measure the amplification of the i-f stage. Express the amplification in decibels.

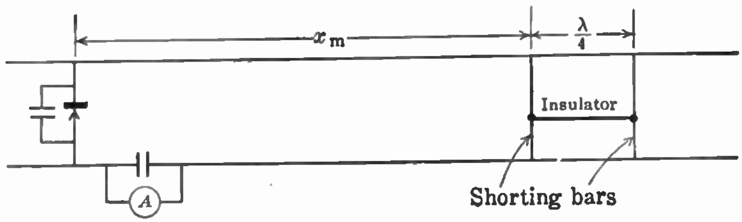
15-9. Negative-Grid Oscillator; Lecher Wires.- The negative-grid u-h-f oscillator is discussed in Ch. 10 and Lecher wires (u-h-f open-wire transmission line) in Ch. 11.

(a) Determine the highest frequency of oscillation of several types of tubes: RCA 955 (use resistor in plate circuit), Hytron 615Y, and RCA 6J5G, or other suitable tubes available. See Fig. 15-10a for a possible oscillator set-up and Fig. 15-10b for a wavemeter arrangement for measuring wavelength. Find the minimum distance x, and the corresponding wavelength, at which each tube continues to oscillate.

(b) Using the 955 tube determine the curve of wavelength λ versus x, and calculate the value of C_t from $C_t = Cx_i$, where C_t is the equivalent capacitance of the tube (under operating conditions) shunting the line, and C is the capacitance per unit length of the open-wire line (Lecher system). The quantity x_i is the intercept with the X-axis of the asymptote to the curve of λ vs x (for x large the curve is approximately a straight line). Consult "Wavelength Characteristics of Coupled Circuits Having Distributed Constants," by Ronald King, Proc. I.R.E., 20,



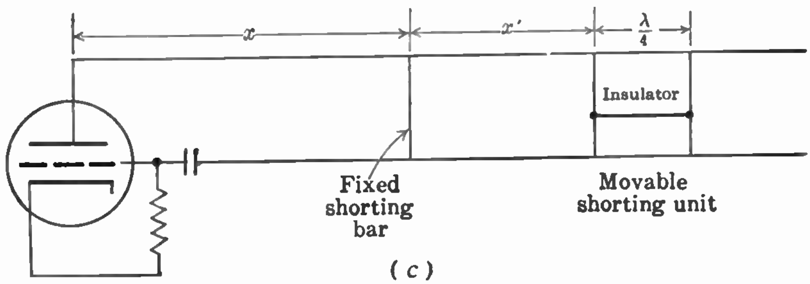
(*a*) Negative-grid oscillator and Lecher wire tuning system.



A = microammeter

For maximum current $x_m = (2n + 1)\frac{\lambda}{2}$

(*b*) Wavemeter (should be coupled very loosely to oscillator).



(*c*)

FIG. 15-10. Lecher - wire uses.

No. 8, 1368, Aug. (1932).

(c) Using the 955 tube, fix a shorting bar at such a position that oscillations of approximately 150-cm wavelength are produced. Using a pair of shorting bars $\lambda/4$ apart, determine the variation of λ with x'. See Fig. 15-10c

(d) Following the method outlined by L. S. Nergaard in "A Survey of Ultra-High-Frequency Measurements," RCA Review, Oct. 1938, determine the resonance wavelength of the 955 tube by plotting $\lambda \tan \beta x$ against λ^2 and taking the square root of the intercept on the λ^2 axis of the straight line obtained for the values of λ^2 not too near the origin.

15-10. Magnetrons and Klystrons.- In Ch. 10 is given a discussion of magnetrons and klystrons.

A. TRANSIT-TIME MAGNETRON OSCILLATORS

Caution: In order to prevent cumulative increase of plate and filament currents, some type of stabilizing circuit must be used with a transit-time magnetron. Either the circuit of Fig. 2-14 or that of Fig. 10-26, or an equivalent circuit must be used. If that of Fig. 2-14 is used, the anode voltage cannot be controlled independently of the plate current. If the circuit of Fig. 10-26 is used, on the other hand, the plate current cannot be controlled independently of the filament current.

(a) Attach the two anode leads to a Lecher wire system 40 or 50 cm in length, provided with suitable radiators at the open end, as shown in Fig. 15-11. Using a crystal detector as

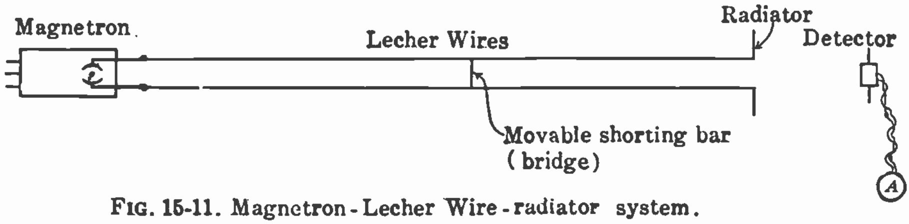


FIG. 15-11. Magnetron-Lecher Wire-radiator system.

an indicator, adjust the field current, angle of the electrode axis relative to the magnetic flux, the filament current, the anode current, and the anode voltage to the various values at which the tube oscillates and the indication of the detector is a maximum. Record the current and voltage readings. With the tube adjusted for maximum amplitude of oscillation, adjust the Lecher wire bridge and record positions at which maximum and minimum energy are radiated. From these readings determine the wavelengths of oscillation. Place the detector near the filament and anode leads and note that high-frequency currents are present.

(b) Introduce quarter-wave concentric-line chokes in the filament and anode leads and repeat (a), noting any differences in behavior, including amplitude of the energy radiated from the end of the Lecher wire system, and the amplitude of high-frequency currents in the filament and anode supply lines.

(c) If a flux meter is available, plot a curve of flux against field current.

(d) Plot curves of wavelength against flux or field current, against anode voltage, and against filament current. Note whether there is any simple relation between the various frequencies of oscillation, and explain the large number of frequencies observed. Explain the effect of the concentric-line chokes. Discuss the curves.

B. KLYSTRON OSCILLATOR

An approximate equivalent circuit for a klystron oscillator is shown in Fig. 15-12.

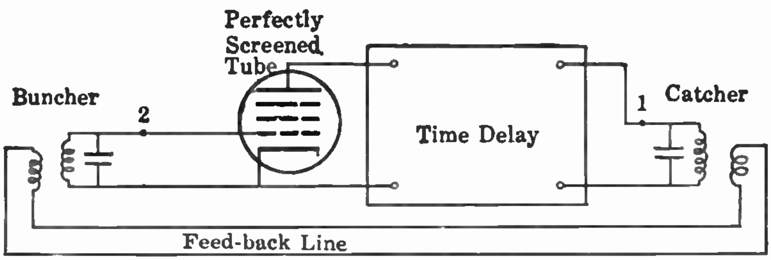


FIG. 15-12. Approximately equivalent circuit of a klystron oscillator.

The "time delay" is the time required by the electrons to travel from buncher to catcher. This time is ordinarily a number of cycles and is a function of the beam voltage applied to the tube.

It is useful to define a transfer impedance $\underline{Z}_{12}$ which is the ratio between the voltage across the buncher (point 2) and the current which causes this voltage, as measured at point 1. This impedance is a complicated function of the frequency, the various circuit constants, some of which may be changed in tuning the klystron, and the feedback line. The phase angle of this impedance may be called φ.

A necessary condition for oscillation is that the phase relations be correct. Specifically, if the circuit be broken at 2 and a small voltage put on the grid, this will cause a plate current which will eventually cause a voltage to appear at the left of the break at 2. For stable oscillation, the voltages on the two sides of the break must be in phase. Analytically, the total phase shift around the circuit must be $2n\pi$ where n is some integer. An equation stating this is [see eq. (10-18b) of Ch. 10]

$$\tau + \varphi = 2n\pi - \frac{\pi}{2} \qquad (10\text{-}17)$$

Here τ is the transit angle in radians or 2π times the time of flight measured in cycles. The extra $\frac{\pi}{2}$ on the right side has to do with the fact that an electron going with mean velocity and in the center of the bunch at the catcher (i.e., the peak of the current wave) is one that passed through the buncher when its voltage was zero. This introduces a quarter-cycle phase shift.

The equation is satisfied by controlling τ by varying the beam voltage, and by the klystron picking an oscillation frequency which gives φ a proper value. The variation of φ with frequency is a somewhat complex matter. However, if a klystron is oscillating with a certain value of φ and a certain beam voltage and associated τ and n, then it can also oscillate in an entirely similar way for a series of other values of voltage: these values correspond to values of τ differing by just 2π. It is the purpose of the experiment to investigate this phenomenon experimentally.

To put the theory in more usable form, let s be the distance between buncher and catcher, v_o the electron velocity, c the velocity of light, f the frequency, and λ the wavelength. Then

$$\tau = 2\pi \frac{s}{v_o} f = 2\pi \frac{s}{\lambda}\frac{c}{v_o}$$

$$= 2\pi \frac{s}{\beta\lambda}$$

where $$\beta = v_o/c$$

The quantity β depends only on the beam voltage and, using known values of e, m, and c it has been stated in Ch. 10 eq. (10-20) that

$$\beta = \sqrt{\frac{V}{250,000}}$$

where V is the beam voltage in volts. Using the value of τ as expressed above,

$$\frac{1}{\beta} = \frac{\lambda}{s}\left(n - \frac{1}{4} - \frac{\varphi}{2\pi}\right) \qquad (10\text{-}20)$$

Thus if the tuning, and so φ, are left alone and the beam voltage is varied to find successive oscillation points, then in the above equation n changes by one unit at a time and if $1/\beta$ is plotted against n, a straight line of slope λ/s should result. It is the object of this experiment to verify the above.

(a) Measure s, the bunching distance. The distance between the first and fourth grids can be measured approximately from the outside of the tube and this, less 0.05", is the grid spacing s.

(b) Connect the klystron output to a crystal and a 10 ma d-c meter. Leave the tuning unchanged and vary the voltages and record them for points of maximum oscillation. Find at least three for each klystron. Compute $1/\beta$ and plot against n.

(c) With the same tuning setting as in (b) measure the wavelength for two klystrons.

(d) Calculate the wavelength from data in (a) and (b) and compare with the results of (c).

(e) By beating the output of the two klystrons together, measure the change in frequency with accelerating voltage and tuning.

(f) Getting n from the value of s in (a), and the accelerating voltage, compute φ and s.

15-11. Impedance Matching with Lumped Parameters.- Impedance matching is important from the point of view of efficiency, maximum power transfer, reduction of reflected waves, etc.

(a) Using the circuit of Fig. 15-14a with the MOPA tightly coupled to the external circuit and with the output in the range of 1.5 to 2.5 megacycles, read the r-f current when the d-c plate current is made a minimum by the tuning condenser. Use a resistance load which makes this minimum current approximately 125 ma (or an appropriate amount for the equipment used).

(b) Measure the open- and short-circuit reactances of a convenient length of transmission line. These values may be obtained by three readings. A length of the order of $\lambda/8$ is suggested.

Reading 1 - Fig. 15-14b. Tune C to resonance, record C.

Reading 2 - Fig. 15-14c. Connect open-circuited coaxial line in parallel with C; retune to resonance and record C. Compute effective X of the open-circuited line (X_{oc}).

Reading 3 - Fig. 15-14d. Connect short-circuited line in series with C. Tune to resonance and compute the effective X of the short-circuited line (X_{sc}).

Compute the characteristic impedance of the line by the relation $\underline{Z} = \sqrt{\underline{Z}_{oc}\underline{Z}_{sc}}$ given in Ch. 11. (Note. If the line has low losses it is not necessary to measure the resistive component of the open- or short-circuit impedances.)

(c) With a small value of the mutual inductance (loose coupling) connect L_2 (Fig. 15-14e) in series with a condenser C_2 and an r-f ammeter. Tune the secondary to resonance. Then connect a resistor R_2 which is equal to the characteristic impedance of the line in series with L_2, C_2 and the r-f ammeter. Be sure that R_2 can dissipate the full output of the MOPA. Increase M until the plate current is increased to the same value as in (a), at the same time making minor adjustments in C_1 to make sure that the tank circuit is tuned to minimum plate current. Compute the power delivered to R_2, which should be nearly the same as that originally delivered to R_1, the load in (a). Read all currents and compare with those obtained in (a).

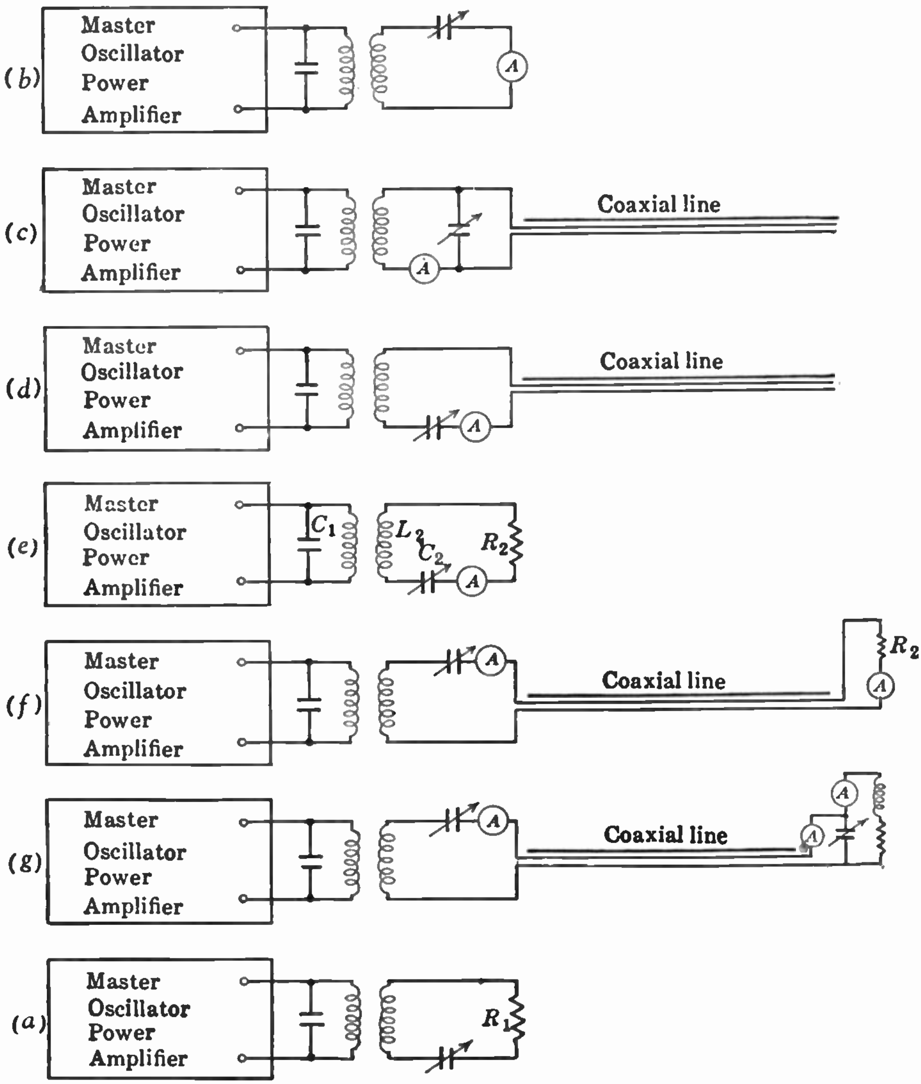


FIG. 15-14. Impedance matching

(d) Replace R_2 of (c) with a line terminated in an r-f meter and R_2. See Fig. 15-14f. Compute the power now delivered to R_2. (Note. If the R_2 used has appreciable reactance some reflection will occur but it is not usually convenient to tune this out.) Read all currents and compare with those obtained previously.

(e) Compute an L matching network (Fig. 15-14g) which can transform the original resistance R_1 into a value equal to R_2. Set up and adjust such a network at the end of the line. Read all currents and compare with those previously obtained. Compute the power delivered to R_1 and the efficiency of the system. Account for the loss of the system by measuring the Q of the coils and estimating the loss in the line.

15-12. Antenna Feeding and Matching with Stubs.- The impedance of an antenna, measured at its input terminals, is not likely to equal the characteristic impedance of the transmission line feeding the antenna. For numerous reasons (efficiency, power transfer, standing waves on the transmission line) it is desirable to match impedances.

(a) Resonant lines. Tune a short-circuited two-wire line to resonance. Measure the current distribution along the line using a crystal loop-type probe. Detune the line by moving the shorting bar a short distance, and retune by means of a variable condenser. Repeat with an open-circuited line. Determine the current distribution along the line when it feeds a half-wave antenna.

(b) Terminate the line with a resistor. Compute the approximate length and location of a stub to be used for proper matching. Attach the stub and adjust the shorting bar for minimum standing waves (Fig. 15-13). Record the ratio of minimum to maximum for each adjustment. Repeat for an open-ended stub.

(c) Repeat (b), using a half-wave antenna in place of the terminating resistor.

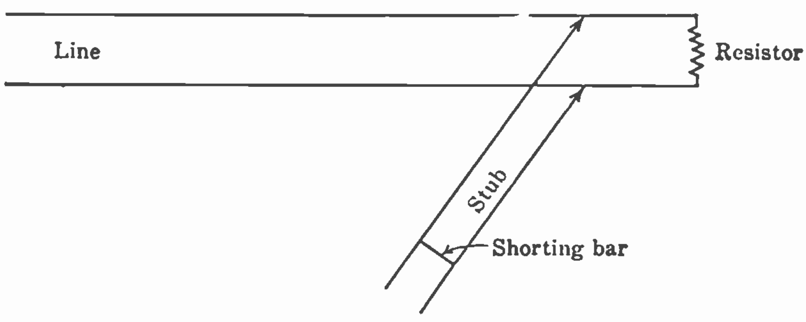


FIG. 15-13. Matching by short-circuited stub.

15-13. Antenna Arrays.- Antennas and antenna arrays are treated in Ch. 12. It is well to perform this experiment in an open space, not in a building

(a) Determine current distribution along a half-wave antenna at a wavelength of approximately 200 cm. Adjust oscillator frequency for maximum current in the antenna. Plot relative antenna current against distance, using a small crystal detector with meter for indicator. Have the antenna well away from metal objects.

(b) Make the measurements of (b) and (c) at approximately 50 cm wavelength. For a measuring device a half-wave antenna with crystal detector at center may be used. A reflector for use with the receiving antenna may prove advantageous. Make measurements with antennas both horizontal and vertical. For feeding the antennas, a coaxial line of variable length (sliding action) may be useful.

Determine radiation patterns of (1) a half-wave antenna; (2) two half-wave antennas, a half-wavelength apart, fed 180° out-of-phase; and (3) two half-wave antennas a half-wavelength apart, fed in phase.

(c) Place a parasitic (unexcited) antenna 0.2 wavelength from a driven antenna. Determine the field pattern for various lengths of the parasitic antenna between 0.35 and 0.7 wavelengths. Note front-to-back ratio.

15-14. Hollow Wave Guides-I.- The latter half of Ch. 14 is given over to a discussion of experiments with hollow-wave guides.

(a) Standing waves in hollow-wave guides. Using the traveling detector obtain data for plotting standing waves when the detector section is terminated (1) with a solid closure; (2) with a small iris opening; (3) with large iris opening; (4) with open end; (5) with a small horn. Terminate with resonant chamber, and attempt to obtain proper termination value with absorbing discs.

(b) Wavelength λ_g in the guide. Using a guide with volume filters, traveling detector, and movable reflector piston adjust the position of the source, filters, and reflector to give standing waves of the type desired and plot probe detector current vs displacement for several types of waves. From these plots determine the wavelength λ_g in the guide for each type of wave.

(c) Using the coaxial wavemeter determine the wavelength in free space.

(d) Using the pinch pipe, determine the curve of transmission against diameter, noting particularly the cutoff diameter.

(e) Effect of bends in guide. Terminate the traveling detector section with a 90° bend in horizontal section, followed by a small horn. Check this arrangement for reflections which will be due to the presence of the bends. Repeat with 90° bend in vertical plane.

(f) Resonance in hollow-wave guides. Attach a resonant chamber with detector to a wave guide, inserting an iris at the junction. Take readings of the detector current and plunger positions over a range sufficient to obtain a good resonance curve. Plot current as a function of plunger position. Repeat with different size irises. Determine maximum and half power positions only. Repeat with (1) polystyrene disc, (2) carbon disc, (3) brass disc.

As an alternative to the above experiment the following one has been found successful and serves to illustrate a number of aspects of u-h-f techniques. It overlaps but does not cover completely the experiment above.

Measurement of Wavelength.

(a) Couple a negative-grid oscillator or other type of ultra-high-frequency oscillator to a section of wave guide, terminated by a movable piston, as shown in Fig. 15-15. (The oscillator may be coupled to the wave guide directly as shown, by means of horns, or by means of a concentric cable.) Couple a doublet radiator to the wave guide by means of a concentric line and matching stubs, as shown in Fig. 15-15. Adjust the piston

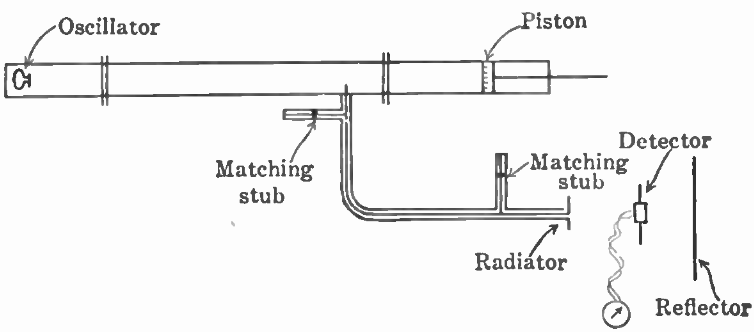


FIG. 15-15. Experimental arrangement for wavelength experiment.

and the matching stubs to give maximum reading with a crystal detector placed at a convenient distance from the radiator. Measure the distances between the settings of one of the matching stubs at which maximum or minimum energy is radiated, and from these readings determine the wavelength in air.

(b) Place a copper reflector beyond the detector in line with the detector and the radiator, as shown in Fig. 15-15, and record positions of the reflector relative to the detector at which the detector gives maximum and minimum readings. From these readings determine the wavelength in air.

(c) Repeat (b), using a stationary reflector and a movable detector.

(d) Record positions of the wave-guide piston at which the radiation is maximum and minimum. From these readings determine the wavelength in the guide. From this value and the dimensions of the guide determine the wavelength in air.

(e) Place the radiator at one end and the detector at the other end of a "pinch pipe" (circular wave guide of adjustable diameter) or "wave press" (rectangular wave guide of adjustable width) and determine the diameter or width at which cutoff occurs. From this reading determine the wavelength in air.

(f) Compare the various values measured and discuss the accuracy of the various methods.

15-15. Hollow Wave Guides-II.- Much pertinent material is discussed in the second half of Ch. 14.

(a) Impedance matching with transfer section. Attach a transfer section to the traveling detector section and feed the energy into a properly terminated coaxial cable. Check for reflections under this condition of operation.

(b) Filter attenuation. Place the filter for transmitting a desired type wave to give maximum transmission. Keeping the power input constant, determine the ratio of desired wave maxima to other maxima. Do this for several types of waves. Using the TE_0, filter, placed as above, plot detector current against angular position of filter using the case when the filter wires are parallel to the feeder.

(c) Field pattern and effect of grid. Terminate the hollow wave guide in a horn of rectangular section, rotating the horn to match the guide to obtain a maximum field strength along the axis. Measure and plot the field for four conditions:

1. Horn open.
2. Wire grid across face of horn and parallel to one side of opening.
3. Wire grid across face of horn and parallel to the other side of opening.
4. Wire grid placed diagonally across opening.

15-16. Horns and Parabolas.- Horns have already been the subject of experimentation in Sec. 15-15. Parabolas are discussed briefly in Ch. 12. The two different types of radiating systems are brought together in this experiment.

(a) Radiation Pattern. Radiate a vertically-polarized 10-cm wave from the horn. Use a vertical half-wave antenna at the focus of the paraboloid for reception. Rotate the parabola about its vertical axis; determine relative response as a function of angular position.

With paraboloid focused on horn, rotate the horn about a vertical axis and determine relative response as a function of angular position.

Rotating horn and antenna so as to produce horizontal polarization, repeat the above.

(b) Check on radiation reciprocity theorem. Driving the dipole in the parabolic reflector and using the horn as a receiver with the same sending power, crystal, and meter, and with the position of the horn and parabola unchanged, repeat and compare the results.

(c) Effects of antenna position in parabaloid. With horn and paraboloid set up as in the first part of (a) vary the position of the antenna along axis, and determine relative response and the field pattern as a function of position along axis of paraboloid. (Keep antenna vertical.)

(d) Polarization effects. Set the antenna at the optimum position (focus of paraboloid), rotate the antenna from the vertical, and determine relative response as a function of angle from vertical.

Chapter 16

A GUIDE TO THE LITERATURE OF ULTRA-HIGH-FREQUENCY TECHNIQUES

Compiled by: Ruth McG. Lane
Vail Librarian, M.I.T.

The purpose of this bibliography is to trace the development of UHF technique from basic background theory to current trends. The compilation is selective only, but includes textbook sources, special publications, and references to technical periodical literature which give a fairly comprehensive survey of the field. After an introductory section, the subject arrangement follows the general outline of the contemporary course in UHF techniques given at engineering schools throughout the U.S. Under each subject, references are arranged chronologically, thus giving a more graphic presentation of developments and at the same time allowing for current additions to keep the bibliography up to date.

The references to periodical literature are of three types: survey type articles which give background information; papers dealing with fundamental theory; and articles which describe recent developments. Those which give also particularly comprehensive references for further reading have been indicated.

U H F: CONTENTS

PERIODICAL TITLE ABBREVIATIONS

A.I.E.E.J. American Institute of Electrical Engineers Journal

A.I.E.E. Trans. American Institute of Electrical Engineers Transactions

Ann. d. Phys. Annalen der Physik

Ann. de Physique Annales de Physique

Bell Labs. Rec. Bell Laboratories Record

B.S.T.J. Bell System Technical Journal

Brown Boveri Rev. Brown Boveri Review

Elec. Com. Electrical Communication

E.E. Electrical Engineering

El. Rev. & W. Elec. Electrical Review and Western Electrician

E.W. Electrical World

Electronic Eng'g Electronic Engineering

E.T.J. Electrotechnical Journal (Japan)

Exp. Wireless Experimental Wireless

G.E. Rev. General Electric Review

Gen. Radio Exp. General Radio Experimenter

Hochfreq. u. Elektroakustik Hochfrequenztechnik und Elektroakustik

I.R.E. Proc. Institute of Radio Engineers Proceedings

I.E.E.J. Institution of Electrical Engineers Journal

I.E.E. of Japan J. Institute of Electrical Engineers of Japan Journal

Japan Electrotech. Lab. Res. Japan Electrotechnical Laboratory Researches

J.A.P. Journal of Applied Physics

J.A.S.A. Journal of Acoustical Society of America

J.F.I. Journal of Franklin Institute

J.O.S.A. Journal of the Optical Society of America

N.B.S. Bul. National Bureau of Standards Bulletin

N.B.S.J. Res. National Bureau of Standards Journal of Research

N.R.C. of Japan Radio Res. National Research Council of Japan

Nippon Elec. Com. Eng'g Nippon Electrical Communication Engineering

Philips Tech. Rev. Philips Technical Review

Phil. Mag. Philosophical Magazine

Phys. Rev. Physical Review

Phys. Soc. (Lond.) Proc. Physical Society (London) Proceedings

Phys. Zeits. Physikalische Zeitschrift

Proc. Roy. Soc. Proceedings of Royal Society

RCA Rev. Radio Corporation of America Review

R.G.E. Revue Générale d'Électricité

R.S.I. Review of Scientific Instruments

Radio Rev. Radio Review

Rev. of Mod. Phys. Reviews of Modern Physics

Soc. Fr. d'Elec. Bul. Société Francaise d'Électricité Bulletin

Tech. Phys. USSR Technical Physics USSR

Wireless Engr. Wireless Engineer

W.W. Radio Rev. Wireless World and Radio Review

Zeits. f. Hochreq. Zeitschrift für Hochfrequenztechnik

Zeits. f. Phys. Zeitschrift für Physik

U H F: INTRODUCTION

Hertz (1887) and Marconi (1922) reported experimental work which visioned the future possibilities of UHF electric waves. The last decade has produced practical demonstrations far beyond their visions and the literature pertaining thereto is extensive. This introductory section presents sources of information which together give a fairly comprehensive summary of UHF developments. Following sections include references to technical details.

Special textbooks:

1. Hertz, Heinrich Electric waves. (Being researches on the propagation of electric action with finite velocity through space.)
2. Hund, August High-frequency measurement, 1933. Phenomena in high-frequency systems, 1936.
3. Hollman, H. E. Physik und Technik der ultrakurzen Wellen, 2 v., 1936. (Extensive bibliography in each volume.)
4. Groos, Otto Einführung in Theorie und Technik der Dezimeterwellen, 1937.
5. Vilbig, Fritz Lehrbuch der Hochfrequenztechnik, Ed. 2, 1939. Schriftumsverzeichnis zum Lehrbuch....1939. (172 pp. Bibliography bound separately.)
6. McIlwain, K. & J. G. Brainerd High-frequency alternating currents, Ed. 2, 1939.
7. RCA Radio at ultra-high frequencies. Technical papers by RCA engineers, 1940.
8. Strutt, M. J. O. Moderne Kurzwellen-Empfängs-Technik, 1939.
9. ----------. Moderne Mehrgitter-Elektronen Rohren, Ed. 2, 1940.
10. Slater, J. C. Microwave Transmission. McGraw, 1942.

Periodical articles:

11. Marconi, G. Radio telephony. I.R.E. Proc. 10:215-38, Aug., 1922.
12. Ultra-short waves. Electrician 109:758, Dec. 7, 1932.
13. Micro-waves. Recent experimental work carried out in Italy by the Marchese Marconi--reception beyond optical range. Electrician 11:3-6, Jan. 6 1933.
14. Clavier, A. G. Production and utilization of micro-rays. Elec. Com. 12:3-11 July, 1933.
15. ---------- & L. C. Gallant. Anglo-French micro-ray link between Lympne and St. Inglevert. Elec. Com. 12:222-28, Jan. 1934.

16. Kolster, F. A. Generation and utilization of ultra-short waves in radio communication.
I.R.E. Proc. 22:1335-53, Dec. 1934.
17. McPherson, W. L. & E. H. Ulrich Micro-ray communication. I.E.E.J. 78:629-57, June 1936. Bibliography: c50 refs. (1865-1935) pp. 656-7. Chronological lists of investigations, p. 657.
18. Ulrich, E. H. Ultra-short wave communication. Elec. Com., 16:64, July 1937.
19. Brillouin, L. Hyperfrequency waves and their practical uses. J.F.I. 229:709-36, June 1940. Elec. Com. 19:4:118-30, 1941.

20. ELECTRONICS: Supplement. April 1942 (Ed. 2 in press)
U-H-F technique
Kandoian, A. G. Radiating systems and wave propagation. pp. 3-8.
Mouromtseff, I. E., R. C. Retherford, and J. H. Findlay Generators for U-H-F waves. pp. 9-14.
Dudley, Beverly U-H-F reception and receivers. pp. 15-19.
Jaffe, D. L. Wide band amplifiers and frequency multiplication. pp. 20-26.
Lewis, R.F. Measurements in the U-H-F spectrum. pp. 27-32.

Progress reports:

21. Institute of Radio Engineers. Annual progress summary.

Radio progress during 1939.
I.R.E. Proc. 28:99-125, Mar. 1940.
Part II. Electronics: UHF tubes. pp. 104-5, 28 refs.
Part III. Wave propagation: UHF pp. 111-12, 17 refs.
22. Radio progress during 1940.
I.R.E. Proc. 29:89-103, Mar. 1941.
Part I. Electronics: UHF tubes, pp. 91-3, 94 refs.
23. Radio progress during 1941.
I.R.E. Proc. 30:57-71, Feb. 1942.
Part I. Electronics: UHF tubes. pp. 60-61, 50 refs.

Bibliography:

24. Kelsey, E. Reference guide to Ultra High Frequencies. [51 pp.] Zenith Radio Corporation, Chicago.

FUNDAMENTALS FOR U H F

Knowledge fundamental to the understanding of UHF techniques includes the theory of amplification, oscillation, modulation and demodulation; the design and use of general purpose electron tubes as circuit elements; and the construction and operation of radio receivers and transmitters. Such information is found in basic textbooks which give references to original sources, and in periodical articles which treat of special problems and developments. The selected list of basic texts given below is supplemented by a short bibliography of leading articles on certain fundamental subjects. For simplicity the latter are arranged as follows:

I. ELECTRONICS: General purpose tubes and circuits

- Fundamental principles
- Tube types, design; operating characteristics; noise in tubes; cathode-ray tubes and circuits (including trigger circuits).

II. OSCILLATORS

III. AMPLIFIERS (including audio, video, radio).

IV. MODULATION (including Frequency modulation) and DEMODULATION, or detection (including Crystal detectors).

V. RECEIVERS and TRANSMITTERS.

Basic textbooks: sources of theoretical information fundamental to UHF techniques. Specific chapter references are cited in various subdivisions of the bibliography, and a few special texts are added.

25. Eastman, A. V. Fundamentals of vacuum tubes. Ed. 2, McGraw, 1941.
26. Everitt, W. L. Communication engineering. Ed. 2, McGraw, 1937.
27. Fink, D. G. Principles of television engineering. McGraw, 1940.
28. Glasgow, R. S. Principles of radio engineering. McGraw, 1936.
29. Henney, K. Principles of radio engineering. Ed. 4, McGraw, 1942.
30. Hoag, J. B. Basic radio. Van Nostrand, 1942.
31. Jordan, E. C. and others. Fundamentals of radio. Prentice-Hall, 1942.
32. Massachusetts Institute of Technology, Electrical Engineering Department Applied electronics. Electron tubes in circuit applications. In press.
33. Millman, J. and S. Seely Electronics. McGraw, 1941.
34. Reich, H. J. Theory and application of electron tubes. McGraw, 1939.
35. Stratton, J. A. Electromagnetic theory. McGraw, 1940.
36. Terman, F. E. Radio engineering. Ed. 2, McGraw, 1937.
37. ----------. Measurements in radio engineering. McGraw, 1935.
38. Zworykin, V. K. and G. A. Morton. Television. Wiley, 1940.

ELECTRONICS: General purpose tubes and circuits.

Textbooks: See Basic Textbooks above. Also the following special texts:

39. MacGregor-Morris, J. T. and J. A. Henley Cathode-ray oscillography. Chapman, 1936.
40. Maloff, I. G. and D. W. Epstein Electron optics in television. McGraw, 1938.
41. Parr, G. Cathode-ray tube and its applications. Ed. 2, Chapman, 1941. Bibliog. 737 refs., pp. 161-76.
42. Rider, J. F. Cathode-ray tube at work. Rider, 1935.
43. Watt, R. A. W. Cathode-ray oscillograph in radio research. H.M.S.O., 1933.

Handbooks and Manuals:

44. RCA Radiotron division Tube handbook. (4v. in 2, loose-leaf)
45. ----------. Manuals: Receiving tube manual. Tech. Ser. RC-14; Cathode-ray tubes and allied types. Tech. Ser. T5-2. Guide for transmitting tubes.
46. ----------. Application notes.
47. General Electric Co. Technical manuals: Receiving tubes; Transmitting tubes.
48. Hygrade Sylvania Corp. Technical manual: Radio tubes.
49. Smith, F. L. Radiotron designer's handbook. Ed. 3, 1940.
50. Western Electric Co. Vacuum tube data sheets.
51. Westinghouse Electric and Manufacturing Co., Special Products Dpt. Radio transmitting tubes. Handbook.

Etc.

Standards:

52. Institute of Radio Engineers Standards: Electronics, 1938. (Reprint, 1942)

Periodical articles:

53. King, R. W. Thermionic vacuum tubes. B.S.T.J. 2:31-100, Oct. 1923.
54. Llewellyn, F. B. Operation of thermionic vacuum tube circuits. B.S.T.J. 5:433-62, July 1926.
55. Kusunose, Y. Calculation of characteristics and design of triodes. Japan. Electrotech. Lab. Research No. 237 (1928) (Condensed version in I.R.E. Proc. 17:1706-49, Oct. 1929.)
56. Hollman, H. E. On the mechanism of electron oscillations in a triode. I.R.E. Proc. 17:229-51, Feb. 1929.
57. Pidgeon, H. A. and J. O. McNally Study of output power obtained from vacuum tubes of different types. I.R.E. Proc., 18:266-93, Feb. 1930.
58. Llewellyn, F. B. Study of noise in vacuum tubes and attached circuits. I.R.E. Proc., 18:243-65, Feb. 1930.
59. Penick, D. B. Measurement and reduction of microphonic noise in vacuum tubes. B.S.T.J., 13:614-33, Oct. 1934.
60. Pearson, G. L. Fluctuation noise in vacuum tubes. B.S.T.J. 13:634-53, Oct. 1934.
61. North, D. O. Analysis of effects of space charge of grid impedance. I.R.E. Proc., 24:108-36, Jan. 1936.
62. Samuel, A. L. Extending the frequency range of the negative grid tube. J.A.P., 8:677-88, Oct. 1937.
63. Reich, H. J. Trigger circuits. Electronics, 12:8:14-17, Aug. 1937. See also,

64. Eccles, W. H. and F. W. Jordan Trigger relay utilizing three-electrode thermionic vacuum tubes. Radio Rev., 1:143-6, Dec. 1919.
65. Hunt, F. V. and R. W. Hickman On electronic voltage stabilizers. R.S.I., 10:6-21, Jan. 1939.
66. American Institute of Electrical Engineers. Joint Subcommittee on Electronics. Decade of progress in electronic means of communication. Prepared by S. B. Ingram. A.I.E.E. Trans., 59:643-9, 1940; Bibliog., 88 refs., pp. 648-9.
67. Thompson, B. J., D. O. North and W. A. Harris Fluctuations in space-charge-limited currents at moderately high frequencies. RCA Rev. 4:269-85, Jan. 1940; 4:441-72, Apr. 1940; 5:106-24, July 1940; 5:244-60, Oct. 1940; 5:371-88, Jan. 1941; 5:505-24, Apr. 1941; 6:114-24, July 1941.
68. Hollman, H. E. UHF oscillography. I.E.E. Proc., 28:213-19, May 1940.
69. North, D. O. and W. R. Ferris Fluctuations induced in vacuum tube grids at high frequencies. I.R.E. Proc., 29:49-50, Feb. 1941.
70. Puckle, O. S. Time bases. I.E.E.J. 89:Part III:100-122, June 1942.
71. Chaffee, E. L. Characteristic curves of the triode. I.R.E. Proc., 30:383-95, Aug. 1942.

OSCILLATORS

Textbooks: (for complete data, see Basic Textbooks, p. 528)

72. Eastman Chap. 11.
73. Everitt Chap. 18.
74. Reich Theory and application . . . Chap. 10.
75. Terman Radio engineering. Chap. 8.

Special texts:

76. Henney, K. Radio engineering handbook. Ed. 3, 1941. Section 9. Vacuum tube oscillators, by R. I. Sarbacher.
77. Prince, D. C. Vacuum tubes as oscillation generators. G.E. Co., 1929.
78. Rider, J. F. Oscillator at work, 1940.
79. Thomas, H. A. Theory and design of valve oscillators, 1939. Bibliog., 83 refs., pp. 261-5.

Periodical articles:

80. Hull, A. W. The dynatron, a vacuum tube possessing negative resistance. I.R.E. Proc., 6:5-37, Feb. 1918.

81. Abraham, H. and E. Bloch Mesure en valeur absolue des periods des oscillations électriques de haute fréquence. Ann. de Physique, 12:237-302, 1919.
82. Prince, D. C. Vacuum tubes as power oscillators. I.R.E. Proc., 11:275-313, 405-35, 527-50, 1923.
83. Horton, J. W. Vacuum tube oscillators. B.S.T.J., 3:508-24, July 1924.
84. Van der Pol, B. On relaxation oscillators. Phil. Mag., 2 (Ser. 7):978-92, Nov. 1926.
85. Pierce, G. W. Magnetostriction oscillators. I.R.E. Proc., 17:42-88, Jan. 1929.
86. Hull, L. M. and J. K. Clapp Convenient method for referring secondary frequency standards to a standard time interval. (Multivibrators) I.R.E. Proc., 17:252-71, Feb. 1929.
87. Scroggiè, M. G. Applications of the dynatron. Wireless Engr., 10:527-40, Oct. 1933.
88. Herold, E. W. Negative resistance and devices for obtaining it. I.R.E. Proc., 23:1201-23, Oct. 1935; Bibliog., 55 refs., pp. 1222-23.
89. LeCorbeiller, P. Non-linear theory of the maintenance of oscillations. I.E.E.J., 79:361-78, Sept. 1936.
90. Pierce, G. W. and A. Noyes, Jr. Improved magnetostriction oscillators. J.A.S.A., 9:185-92, Jan. 1938.
91. Scott, H. H. New type of selective circuit and some applications. (RC oscillators) I.R.E. Proc., 26:226-36, Feb. 1938.
92. Meacham, L. A. Bridge stabilized oscillatòr. B.S.T.J., 17:574-91, Oct. 1938; also in I.R.E. Proc., 26:1278-94, Oct. 1938.
93. Ginzton, E. L. and L. M. Hollingsworth Phase-shift oscillators. I.R.E. Proc., 29:43-9, Feb. 1941.
94. DeLaup, P. S. Sine-waves in RC oscillators. Electronics, 14:34-6, Jan. 1941.

AMPLIFIERS

Textbooks: (for complete data, see Basic Textbooks, p. 528)

95. Chaffee Chap. 12-17.
96. Eastman Chap. 9-10.
97. Everitt Chap. 15-17.
98. Fink Chap. 6.
99. Glasgow Chap. 7, 9-10.
100. Reich Theory and applications . . . Chap. 6-9.
101. Terman Radio engineering, Chap. 5-7.
102. Zworykin and Morton Chap. 14.
103. Henney Radio engineering handbook; Section 11. Audio-frequency amplifiers, G. Koehler; Section 12. Radio-frequency amplifiers, R. S. Glasgow.

Periodical articles:

104. DeForest, Lee Audion, a new receiver for wireless telegraphy. A.I.E.E. Proc., 25:719-47, Oct. 1906.
105. ----------. Audion as amplifier and detector. I.R.E. Proc., 2:15-29, Jan. 1914.
106. Van der Bijl, H. J. Theory and operating characteristics of the thermionic amplifier. I.R.E. Proc., 7:97-128, Apr. 1919.
107. Carson, J. R. Theoretical study of the three-element vacuum tube. I.R.E. Proc., 7:187-200, Apr. 1919.
108. Armstrong, E. H. New method of short-wave amplification. I.R.E. Proc., 9:3-27, Feb. 1921.
109. Spitzer, E. E. Grid losses in power amplifiers. I.R.E. Proc., 17:985-1005, June 1929.
110. Fay, C. E. Operation of vacuum tubes as Class B and C amplifiers. I.R.E. Proc., 20:548-68, Mar. 1932.
111. Luck, D. G. C. Resistance-coupled amplifiers. I.R.E. Proc., 20:1401-6, Aug. 1932.
112. Nyquist, H. Regeneration theory. B.S.T.J., 11:126-47, Jan. 1932.
113. Thompson, B. J. Graphical determination of performance of push-pull audio amplifiers. I.R.E. Proc., 21-591-600, Apr. 1933.
114. Black, H. S. Stabilized feed-back amplifier. B.S.T.J., 13:1-18, Jan. 1934; E.E., 53:114-20, Jan. 1934.
115. Everitt, W. L. Optimum operating characteristics for Class C amplifiers. I.R.E. Proc., 22:152-76, Feb. 1934.
116. Robinson, H. A. Operation of R-F power amplifiers. QST, Feb. 1934, pp. 25-8; Apr. 1934, pp. 14-18, 76-80.
117. Llewellyn, F. B. Phase angle of vacuum tube transconductance at very high frequencies. I.R.E. Proc., 22:947-56, Aug. 1934.
118. Johnson, J. B. and F. B. Llewellyn Limits to amplification. E.E., 53:1449-54, Nov. 1934; B.S.T.J., 14:85-96, Jan. 1935.
119. Mouromtseff, I. E. and H. N. Kozanowski Analysis of operation of vacuum tubes as Class C amplifiers. I.R.E. Proc., 23:752-78, July 1935. Supplementary notes. I.R.E. Proc., 24:654-6, Apr. 1936.
120. Terman, F. E. and W. C. Roake Calculation and design of Class C amplifiers. I.R.E. Proc., 24:620-32, Apr. 1936.
121. Ferris, W. R. Input resistance of vacuum tubes as UHF amplifiers. I.R.E. Proc., 24:82-107, Jan. 1936.
122. Samuel, A. L. and N. E Sowers Power amplifier for ultra-high Frequencies. I.R.E. Proc., 24:1464-83, Nov. 1936.

123. Seeley, S. W. and C. N. Kimball Analysis and design of video amplifiers. RCA Rev., 2:171-83, Oct. 1937; 3:290-308, Jan. 1939.
124. Ginzton, E. L. Balanced feed-back amplifiers. I.R.E. Proc., 26:1367-79, Nov. 1938.
125. Strutt, M. J. O. and A. van der Ziel Causes of increase of the admittances of modern high-frequency amplifier tubes on short waves. I.R.E. Proc., 26:1011-32, Aug. 1932; Bibliog., 30 refs., pp. 1131-2.
126. Everest, F. A. Wide band television amplifiers. Electronics, 11:1:16-19, Jan. 1938; 11:5:24-7, May 1938.
127. Chaffee, J. G. Application of negative feed-back to FM. I.R.E. Proc., 27:317-31, May 1939; B.S.T.J., 18:404-37, July 1939.
128. Carson, J. R. FM - theory of the feedback circuit. B.S.T.J., 18:395-403, July 1939.
129. Bedford, A. V. and G. L. Fredendall Transient response of multi-stage video-frequency amplifiers. I.R.E. Proc. 27:277-84, Apr. 1939.
130. Wheeler, H. A. Interpretation of amplitude and phase distortion in terms of paired echoes. I.R.E. Proc., 27:359-84, June 1939.
131. ----------. Wide-band amplifiers for television. I.R.E. Proc., 27:429-38, July 1939.
132. Swift, G. Amplifier testing by means of square waves. Communications 19:2:22-6,52, Feb. 1939.
See also,
133. Arguimbau, L. B. Network testing with square waves. Gen. Radio Exp., Dec. 1939.
134. Terman, F. E. Some applications of negative feedback (with special reference to laboratory equipment). I.R.E. Proc., 27:649-55, Oct. 1939.
135. Strutt, M. J. O. and A. van der Ziel New push-pull amplifier for decimeter waves. Philips Tech. Rev., 5:172-81, June 1940.
136. Rodwin, G. High-gain amplifier for 150 megacycles. I.R.E. Proc., 28:257-61, June 1940.
137. Foster, D. E. and J. A. Rankin Video output systems. RCA Rev., 5:409-38, Apr. 1941.
138. Jaffe, D. L. Wide-band amplifiers and frequency multiplication. Electronics, 15:4:56-62, Apr. 1942.

MODULATION and DEMODULATION

Textbooks: (For complete data, see Basic Textbooks, p. 528)

139. Everitt Chap. 13-14.
140. Glasgow Chap. 11-12.

141. McIlwain and Brainerd Chap. 6-7.
142. Terman: Radio engineering Chap. 9-10.
143. Rider, J. F. FM: an introduction to frequency modulation, 1940.

Periodical articles:

144. Heising, R. A. Modulation in radio telephony. I.R.E. Proc., 9:305-52, Aug. 1921.
145. Carson, J. R. Notes on the theory of modulation. I.R.E. Proc., 10:57-64, Feb. 1922.
146. Hartley, R. V. L. Relation of carrier and side bands in radio transmission. B.S.T.J., 2:90-112, Apr. 1923.
147. Van der Pol, B. Frequency modulation. I.R.E. Proc., 18:1194-1205, July 1930.
148. Roder, H. Amplitude, phase, and frequency modulation. I.R.E. Proc., 19:2145-76, Dec. 1931. Discussion ... by D. G. C. Luck, I.R.E. Proc., 20:884-7, May 1932
149. Carrara, N. Detection of microwaves. I.R.E Proc., 20:1615-25, Oct. 1932.
150. Kilgour, C. E. and J. M. Glessner Diode detection analysis. I.R.E. Proc., 21:930-43, July 1933.
151. Chaffee, J. G. Determination of frequency-modulated waves. I.R.E. Proc., 23:517-40, May 1935.
152. Armstrong, E. H. New method of reducing disturbances in radio signalling by a system of frequency modulation. I.R.E. Proc., 24:689-740, May 1936.
153. Carson, J. R. and T. C. Fry Variable frequency electric circuit theory with application to the theory of frequency modulation. B.S.T.J., 16:513-40, Oct. 1937.
154. Crosby, M. G. Frequency modulation noise characteristics. I.R.E. Proc., 25:472-514, Apr. 1937.
155. ----------. Carrier and side-frequency relations with multitone frequency or phase modulation. RCA Rev., 3:103-6, July 1938.
156. Weir, I.R. Field tests of frequency and amplitude modulation with ultra-high-frequency waves. G.E. Rev., 42:188-91; 270-73, May-June 1939.
157. Everitt, W. L. Frequency modulation. A.I.E.E. Trans. 59:613-24, 1940.
158. Guy, R. F. and R. M. Morris NBC frequency modulation field test. RCA Rev., 5:190-225, Oct. 1940.
159. Seeley, S. W. Frequency modulation. RCA Rev., 468-80, Apr. 1941; Bibliog., 39 refs., pp. 478-80.
160. ----------. Generation and detection of FM waves. RCA Rev., 6:269-86, Jan. 1942.
161. Wheeler, H. A. Common-channel interference between two FM signals. I.R.E. Proc., 30:34-50, Jan. 1942.

162. Rettenmeyer, F. X. Radio-electronic bibliography 2. Frequency modulation. Radio, June 1942, pp. 35-38.

DEMODULATION with special reference to CRYSTAL DETECTORS

Textbooks:

163. Pierce, G. W. Principles of wireless telegraphy, 1910. Chap. 17-18, Crystal rectifiers.

164. Blake, G. G. History of radio telegraphy and telephony, 1926, Chap. 7, Electrolytic and crystal detectors.

165. Jolley, L. B. W. Alternating-current rectification. Ed. 2, 1926, Chap. 17, Crystal rectifiers; Bibliog., pp. 423-4.

166. Morecroft, J. H. Radio communication. Ed. 3, 1933, pp. 441-6.

Periodical articles:

167. Pickard, G. W. Thermo-electric wave detectors. (Letter to Editor) E W, 48:1003, 24, Nov. 1906.

168. Pierce, G. W. Crystal rectifiers for electric currents and electric oscillations. I. Carborundum., Phys. Rev., 25:31-60, July 1907; II. Carborundum, molybdenite, anatase, brookite. Phys. Rev., 28:153-87, Mar. 1909; III. Iron pyrites, Phys. Rev., 29:478-84, Nov. 1909.

169. Flowers, A. E. Crystal and solid contact rectifiers. Phys. Rev., 29:445-60, Nov. 1909.

170. ----------. Characteristics of crystal rectification. Phys. Rev. (Ser. 2) 3:25-46, Jan. 1914.

171. Eccles, W. H. Electrothermal phenomena at the contact of two conductors, with a theory of a class of radiotelegraph detectors. Phys. Soc. (Lond.) Proc., 25:273-93, June 1913.

172. Austin, L. W. High resistance contact thermo-electric detector for electric waves. Phys. Rev., 24:508-10, June 1907.

173. ----------. Some contact rectifiers of electric currents. N.B.S. Bul., 5:133-47, Aug. 1908.

174. ----------. Comparative sensitiveness of some common detectors of electrical oscillations. N.B.S. Bul., 6:527-42, Nov. 1910.

175. Hunt, V. A. and L. E. Whittemore Some characteristics of crystal detectors. Phys. Rev. (Ser. 2) 7:518-28, May 1916.

176. Strachan, J. Crystal detection in theory and practice. W.W. and Radio Rev., 14:216-19, 245-9, 289-81, 396-8, 424-6, 462-4, 21 May-16 July 1924.

177. Gabel, V. Crystal as generator and amplifier. W.W. and Radio Rev., 15:2-5, 47-50, 1 Oct.-8 Oct. 1924.
178. Colebrook, F. M. Rectifying detectors. Exp. Wireless, 2:330-40, 394-403, 459-68, Mar.-May 1925.
179. Ogawa, W. Analogy between crystal detector and vacuum detector. Phil. Mag. (Ser. 7) 6:175-8, July 1928.

RECEIVERS and TRANSMITTERS

Textbooks: (For complete data, see Basic Textbooks, p. 528)

180. Glasgow Chap. 13 Receiving systems.
181. Terman: Radio engineering Chap. 12 Transmitters; Chap. 13, Radio receivers.
182. A.R.R.L. Radio amateur's handbook. Ed. 19, 1941.
183. "Radio" handbook. Ed. 8, 1941.
184. Henney, K. Radio engineering handbook. Ed. 3, 1941.

Standards: Institute of Radio Engineers

185. Standards on radio receivers. 1938 (1942 reprint)
186. Standards on transmitters and antennas. 1938 (1942 reprint)

Periodical articles:

187. Armstrong, E. H. Some recent developments in regenerative circuits. I.R.E. Proc., 10:244-60, Aug. 1922.
188. Byrnes, I. F. and J. B. Coleman 20-40 kilowatt high-frequency transmitter. I.R.E. Proc., 18:422-49, Mar. 1930.
189. Llewellyn, F. B. Rapid method of estimating signal-to-noise ratio of a high-gain receiver. I.R.E. Proc., 19:416-21, Mar. 1931.
190. Labus, J. W. and H. Roder Suppression of radio-frequency harmonics in transmitters. I.R.E. Proc., 19:949-62, June 1931.
191. Ballantine, S. High quality radio broadcast transmission and reception. I.R.E. Proc., 22:564-629, May 1934; 23:618-652, June 1935.
192. Chambers, J. A. and others WLW 500-kw. broadcast transmitter. I.R.E. Proc., 22:1151-80, Oct. 1934.
193. Wolff, I., E. G. Linder and R. A. Braden Transmission and reception of centimeter waves. I.R.E. Proc., 23:11-23, Jan. 1935.
194. Lindenblad, N. E. Development of transmitters for frequencies above 300 megacycles. I.R.E. Proc., 23:1013-Sept. 1935.
195. Hull, R. A. New receiving system for ultra-high frequencies. QST 19:Nov. 1935:10-14; 19:Dec. 1935:31-7.

196. Staka, H. Superregeneration of an ultra-short-wave receiver. I.R.E. Proc., 23:841-84, Aug. 1935.
197. Dunmore, F. W. Unicontrol radio receiver for ultra-high frequencies using concentric lines as interstage couplers. I.R.E. Proc., 24:837-49, June 1936.
198. Scroggie. B. G. Superregenerative receiver. Wireless Engr., 13:581-92, Nov. 1936.
199. Zworykin, V K., G. A. Morton and L. Malter Secondary emission multiplier. I.R.E. Proc., 24:351-75, Mar. 1936.
200. Mayer, H. F. Automatic selectivity control. Electronics, 9:12:32-4, Dec. 1936.
201. Benjamin, M., C. W. Cosgrove and G. W. Warren Modern receiving valves: design and manufacture. I.E.E.J., 80:401-39, Apr. 1937; Bibliog., 40 refs., p. 431.
202. Reeves, A. H. and E. H. Ullrich Superheterodyne reception of microwaves. Elec. Com., 16:153-7, Oct. 1937.
203. Frink, F. W. Basic principles of superregenerative reception. I.R.E. Proc., 26:76-106, Jan. 1938.
204. Picker, W. J. (Development of wireless transmitting valves.) Chairman's address: Wireless Section, I.E.E., I.E.E.J., 88:Part I:38-46, Jan. 1941.
205. DeWalt, K. C. Three new ultra-high-frequency triodes. I.R.E. Proc., 29:475-80, Sept. 1941.
206. Thompson, B. J. Voltage-controlled electron multipliers. I.R.E. Proc., 29:583-7, Nov. 1941.
207. Wagner, H. W. and W. R. Ferris Orbital-beam secondary-electron multiplier for ultra-high-frequency amplification. I.R.E. Proc., 29:598-602, Nov. 1941.
208. Haller, C. E. Design and development of three new ultra-high frequency transmitting tubes. I.R.E. Proc., 30:20-26, Jan. 1942.
209. Herold, E. W. Analysis of signal-to-noise ratio of ultra-high frequency receivers. RCA Rev., 6:302-31, Jan. 1942.
210. ----------. Operation of frequency converters and mixers for superheterodyne reception. I.R.E. Proc., 30:84-103, Feb. 1942.
211. North, D. O. Absolute sensitivity of radio receivers. RCA Rev., 6:332-43, Jan. 1942.
212. Dudley, B. UHF reception and receivers. Electronics, 15:4:51-5, April 1942; Bibliog., 45 refs., p. 55.

U H F TECHNIQUES

To understand the theory and practice of UHF techniques, which differ greatly from those of radio communication at lower frequencies, special knowledge is necessary. The use of ultra-high-frequency waves requires special adaptation of the theory of circuits and wave phenomena; constructional features and the physical appearance of circuit elements differ. The following bibliography offers a selection of references from the extensive UHF literature which has appeared within a decade. The references, chiefly to periodical articles, are grouped under topics as follows:

UHF TECHNIQUES

- GENERATORS
  - General
  - Negative-grid tubes
  - Barkhausen oscillators
  - Velocity modulation tubes
    - Klystron
  - Magnetrons
- TRANSMISSION
- RADIATION
  - Antennas and arrays
  - Electromagnetic horns
- WAVE GUIDES
  - Cavity resonators
- PROPAGATION

Periodical articles:

213. Hertz, H. Uber sehr schnelle electrische Schwingungen. Wiedermann's Annalen, 31:421-48, 1887; also in Hertz, H. Electric waves, 1893, Chap. 2.

214. Benham, W. E. Theory of the internal action of thermionic systems at moderately high frequencies. I. Phil. Mag. (7) 5:641-62, Mar. 1928; II. Phil. Mag. (7) 11:457-517, Feb. 1931.

215. Rohde, L. Uber senderöhren zur erzeugung von meterwellen. Hochfreq.u.Elektroakustik, 40:3-5, July 1932.

216. White, W. C. Selection of types of vacuum tubes as sources of high-frequency power. G. E. Rev., 36:394-7, Sept. 1933.

217. Llewellyn, F. B. Vacuum tube electronics at ultra-high frequencies. I.R.E. Proc. 21:153-73, Nov. 1933; B.S.T.J. 13:59-101, Jan. 1934; Bell Mon. B-770.

218. Salzberg, B. Design and use of "acorn" tubes for ultra-high frequencies. Electronics 7:282-3,293, Sept. 1934.

219. Benham, W. E. Mechanism of electronic oscillations. Wireless Engr. 12:3-7, Jan. 1935.

220. Llewellyn, F.B. Note on vacuum tube electronics at ultra-high frequencies. I.R.E.Proc. 23:112-27, Feb. 1935; Bell Mon. B-848.

221. ----------. Operation of ultra-high-frequency vacuum tubes. B.S.T.J. 14:632-65, Oct. 1935; Bell Mon. B-890.

222. Salzberg, B. Recent developments in miniature tubes. I.R.E.Proc. 23:1142-57, Oct. 1935.

223. Benham, W. E. Contribution to tube and amplifier theory. I.R.E.Proc. 26:1093-1170, Sept. 1938, (copious foot-note refs.)

224. Schade, O. H. Beam power tubes. I.R.E.Proc. 26:137-81, Feb. 1938.

225. Thompson, B. J. Review of ultra-high-frequency vacuum tube problems. RCA Rev. 3:146-55, Oct. 1938.

226. Black, D. H. Ultra-short wave oscillators. Elec. Com. 17:325-34, Apr. 1939.

227. Llewellyn, F. B. Production of ultra-high-frequency os-cillations by means of diodes. B.S.T.J. 18:280-91, Apr. 1939.

228. King, R. Beam tubes as ultra-high-frequency generators. J.A.P. 10:638-47, Sept. 1939.

229. De Walt, K. C. Three new ultra-high-frequency triodes. I.R.E.Proc. 29:475-80, Sept. 1941.

230. Lüdi, F. Generators for ultra-short waves. Brown Boveri Rev. 28:393-6, Dec. 1941.
231. Mouromtseff, I.E. and others. Ultra-high-frequency generators. Electronics 15:4;45-50, Apr. 1942; Bibliog. 25 refs. p. 50.

U H F GENERATORS: Negative-grid oscillators

Periodical articles:

232. White, W. C. Pliotron oscillator for extreme frequencies. G.E. Rev. 19:771-5, 1916.
233. Van der Pol, B. Production and measurement of short continuous electromagnetic waves. Phil. Mag. (6) 38:90-97, 1919.
234. Southworth, G. C. Electron tube generators of alternating currents of ultra-radio frequencies. Radio Rev. 1:577-84, 1920.
235. McArthur, E. P. and E. E. Spitzer. Vacuum tubes as high-frequency oscillators. I.R.E.Proc. 19:1971-82, Nov. 1931; Bibliog. 28 refs.
236. Wenstrom, W. H. Experimental study of regenerative ultra-short-wave oscillators. I.R.E.Proc. 20:113-20, Jan. 1932.
237. Mouromtseff, I. E. and H. V. Noble. New type of ultra-short-wave oscillator. I.R.E.Proc. 20:1328-44, Aug. 1932.
238. Thompson, B. J. and G. W. Rose. Vacuum tubes of small dimensions for use at extremely high frequencies. I.R.E.Proc. 21:1707-21, Dec. 1933; also in RCA : Radio at UHF, 1940, pp. 334-48.
239. Fay, C. E. and A. L. Samuel. Vacuum tubes for generating frequencies above 100 megacycles. I.R.E.Proc. 23:199-212, Mar. 1935; Bibliog. 15 refs., p. 212; Bell Mon. B-855.
240. Llewellyn, F. B. Equivalent networks of negative-grid vacuum tubes at ultra-high frequencies. B.S.T.J. 15:575-86, Oct. 1936.
241. Samuel, A. L. Extending the frequency range of the negative-grid tube. J.A.P. 8:677-88, 1937.
242. ----------. Negative-grid triode oscillator and amplifier for ultra-high frequencies. I.R.E.Proc. 25:1243-52, Oct. 1937; Bell Mon. B-1027.
243. Wagener, W. G. Requirements of a new ultra-high frequency power tube. RCA Rev. 2:258-65, Oct. 1937.
244. ----------. Developmental problems and operating characteristics of two new ultra-high frequency triodes. I.R.E.Proc. 26:401-414, Apr. 1938.

245. Barrow, W. L. Oscillator for ultra-high frequencies. R.S.I. 9:170-74, June 1938.

246. Lockhart, C. E. Generation and amplification of microwaves; (Negative-grid valves). I. Electronic Eng'g 14:336-8,347, Aug. 1941; II. Electronic Eng'g 14:384-7,414, Sept. 1941; Bibliog. 49 refs. p. 414; III. Electronic Eng'g 14:432-4, Oct. 1941; IV. Electronic Eng'g 14:530-33, Dec. 1941.

U H F GENERATORS: Barkhausen oscillators

Periodical articles:

247. Barkhausen, H. and K. Kurz. Die Kürzesten mit Vakuumröhren herstellenbaren Wellen. Phys. Zeits. 21:1-6, Jan. 1920.

248. Gill, E. W. B. and J. H. Morrell. Short electric waves obtained by valves. Phil. Mag. 44:161-78, 1922.

249. ----------. Short electric waves obtained by use of secondary emission. Phil. Mag. 49:369-79, 1925.

250. Hollman, H. E. On the mechanism of electron oscillations in a triode. I.R.E.Proc. 17:229-51, Feb. 1929 (37 refs).

251. Gill, E. W. B. Electrical oscillations of very short wavelengths. Phil. Mag. 12:843-53, Oct. 1931.

252. Potapenko, G. Investigations in the field of ultra-short electromagnetic waves. I. Generator for production of ultra-short undamped waves; II. Normal and dwarfed waves. Phys. Rev. 39:625-65 (47 refs.), 1932; III. Electronic oscillations in vacuum tubes and causes of generation of dwarfed waves. Phys. Rev. 40:988-1001 (22 refs.), 1932; IV. On the dependence of ultra-short electromagnetic waves upon the heating current and upon the amplitude of oscillations. Phys. Rev. 41:216-30 (15 refs.), 1932.

253. Kozanowski, H. N. New circuit for production of ultra-short-wave oscillations. I.R.E.Proc. 20:957-68, June 1932.

254. Potapenko, G. On the generation of ultra-short electromagnetic waves using the scheme of H. Barkhausen and K. Kurz. Phil. Mag. 14:1127-42, Dec. 1932.

255. Megaw, E. C. S. Electronic oscillations. I.E.E.J. 72:313-25, Apr. 1933; Bibliog. 47 refs.; Discussion, pp. 348-52.

256. Thompson, B. J. and P. D. Zottu. Electron oscillator with plane electrodes. I.R.E.Proc. 22:1374-85, Dec. 1934.

257. Kelly, M. J. and A. L. Samuel. Vacuum tubes as high-frequency oscillators. E.E. 53:1504-17, Nov. 1934; also in B.S.T.J. 14:97-134, Jan. 1935; Bell Mon. B-839.

U H F GENERATORS: Velocity modulated tubes

Periodical articles:

258. Heil, A. A. and O. Heil. Eine neue Methode zur Erzeugung kurzer, ungedämfter elektromagnetischer Wellen grosser Intezität. Zeits.f.Phys. 95:752-62, (1935).
259. Brüche, E. and A. Recknagel. Über die "Phasenfokussierung" bei der Elektronenbewegung in schnellveränderlichen elektrischen Felden. Zeits.f.Phys. 108:459-82, (1938).
260. Hansen, W. W. A type of electrical resonator. J.A.P. 9:654-63, Oct. 1938.
261. Hahn, W. C. and G. F. Metcalf. Velocity-modulated tubes. I.R.E.Proc. 27:106-16, Feb. 1939.
262. Haeff, A. V. Ultra-high-frequency power amplifier of novel design. Electronics 12:30-32, Feb. 1939.
263. Varian, R. H. and S. H. Varian. High-frequency amplifier and oscillator. J.A.P. 10:140, Feb. 1939.
264. Hansen, W. W. and R. D. Richtmyer. On resonators suitable for klystron oscillators. J.A.P. 10:189-99, Mar. 1939.
265. Varian, R. H. and S. F. Varian. High-frequency oscillator and amplifier. J.A.P. 10:321-7, May 1939.
266. Hahn, W. C. Small signal theory of velocity-modulated electron beams. G.E.Rev. 42:258-70, June 1939.
267. Haeff, A. V. Effect of electron transit time on the efficiency of a power amplifier. R.C.A.Rev. 3:114-22, July 1939.
268. Hahn, W. C. Wave energy and transconductance of velocity-modulated electron beams. G.E.Rev. 42:497-502, Nov. 1939.
269. Webster, D. L. Cathode-ray bunching. J.A.P. 10:501-8, July 1939.
270. Ramo, S. Space charge and field waves in an electron beam. Phys. Rev. 56:276-83, 1 Aug. 1939.
271. Webster, D. L. Theory of klystron oscillations J.A.P. 10:864-72, Dec. 1939.
272. Ramo, S. Electronic-wave theory of velocity modulation tubes. I.R.E.Proc. 27:757-63, Dec. 1939.
273. Condon, E. U. Electronic generation of electromagnetic oscillations. J.A.P. 11:502-6, July 1940.
274. Ramo, S. Traveling waves in electron beams. Communications 20:5-8, Nov. 1940.

275. Kompfner, R. Velocity modulation. Wireless Engr. 17:472-88, Nov. 1940; Bibliography, 18 refs., p. 488.

276. Benham, W. E. Phase focussing in velocity modulated beams. Wireless Engr. 17:514-16, Dec. 1940; Bibliography, 14 refs., p. 516

277. Kompfner, R. Transit-time phenomena in electronic tubes. Wireless Engr. 19:2-7, Jan. 1942.

U H F GENERATORS: Magnetrons

M.I.T.E.E. Dpt. Thesis

278. McDonald, D. Three-element magnetron. B.S., 1941; Bibliog. 394 refs.

Periodical articles:

279. Hull, A. W. Effect of a uniform magnetic field on the motion of electrons between coaxial cylinders. Phys. Rev. (2) 18:31-57, 1921.

280. ----------. The Magnetron. A.I.E.E.J. 40:715-23, Sept. 1921.

281. ----------. Axially controlled magnetron. A.I.E.E. Trans. 42:915-20, 1923.

282. Elder, F. R. Magnetron amplifier and oscillator. I.R.E. Proc. 13:159-88, Apr. 1925.

283. Habann, E. Eine neue Generatorröhre. Zeits.f.Hochfreq. 24:115-20, 1924

284. Okabe, K. Undamped extra-short electromagnetic waves obtained with magnetron. I.E.E. of Japan J. 469:99-101, 1927.

285. ----------. Production of intense extra-short electromagnetic waves with "split-anode magnetron." I.E.E. of Japan J. 476:33-4, 1928.

286. ----------. On the short-wave limit of magnetron oscillators. I.R.E.Proc. 17:652-9, 1929.

287. ----------. On the magnetron oscillation of a new type. I.R.E.Proc. 18:1748-9, 1930.

288. Kilgore, G. R. Magnetostriction oscillators for generation of ultra-short waves. I.R.E.Proc. 20:1741-51, Nov. 1932.

289. Megaw, E. C. S. Investigation of magnetron short-wave oscillator. I.E.E.J. 72:326-48. 1933.

290. Posthumus, K. Principles underlying generation of oscillations by means of split-anode magnetron. Philips Transmitting News 1:3:11-25, Dec. 1934.

291. Okabe, K. On the production of ultra-short wave oscillations with double-anode magnetron. N.R.C. of Japan Radio Res. 5:69-75, 1935.

292. Linder, E. G. Description and characteristics of the end-plate magnetron. I.R.E.Proc. 24:633-53, Apr. 1936.
293. Kilgore, G. R. Magnetron oscillator for generation of frequencies between 300 and 600 megacycles. I.R.E. Proc. 24:1140-57, Aug. 1936; Bibliog. 27 refs. (footnotes); also in RCA: U H F, 1940, pp. 360-77.
294. Herriger, F. u. F. Hülster. Die Schwingungen der Magnetfeldröhren. Hochfreq.u.Elektroakustik 49:123-32, Apr. 1937; (Corrections: July 1937, p. 35).
295. Kilgore, G. R. Magnetron as high-frequency generator. J.A.P. 8:666-76, Oct. 1937; Bibliog. 34 refs., p. 676.
296. Okabe, K. Mechanism of split-anode magnetron. Nippon Elec. Com. Eng'g, Dec. 1937, pp. 530-35.
297. Linder, E. G. Effect of high energy electron random motion upon the shape of magnetron cut-off curve. J.A.P. 9:331-4, May 1938.
298. Heller, G. Magnetron as generator of ultra short waves. Philips Tech. Rev. 4:189-97, July 1939.
299. Linder, E. G. Anode-tank-circuit magnetron. I.R.E.Proc. 27:732-8, Nov. 1939.
300. Omoto, Y. and K. Morita. Electrostatic field and capacitance between split anodes of a magnetron. E.T.J. 4:147-50, July 1940.
301. Brillouin, L. Theory of the magnetron. Elec. Com. 20: 112-21, 1941.

U H F TRANSMISSION

Periodical articles:

302. Espenschied, L. Application to radio of wire transmission engineering. B.S.T.J. 1:117-41, Nov. 1922.
303. Gannett, D. K. and E. I. Green. Wire transmission system for television. B.S.T.J. 6:616-32, Oct. 1927.
304. Carson, J. R. Rigorous and approximate theories of electrical transmission along wires. B.S.T.J. 7:11-25, Jan. 1928.
305. Green, E. I. Transmission characteristics of open-wire lines at carrier frequencies. B.S.T.J. 9:730-59, Oct. 1930.
306. Conklin, J. W., J. L. Finch and C. W. Hansell. New methods of frequency control employing long lines. I.R.E.Proc. 19:1918-30, Nov. 1931.
307. Sterba, E. J. and C. B. Feldman. Transmission lines for short-wave radio systems. B.S.T.J. 11:411-50, July 1932; also in I.R.E.Proc. 20:1163-1202, July 1932; Bell Mon. B-683.

308. Terman, F. E. Resonant lines in radio circuits. E.E. 53:1046-53, July 1934.

309. Espenschied, L. and M. E. Strieby. Systems for wide-band transmission over coaxial lines. E.E. 53:1371-80, Oct. 1934; also in B.S.T.J. 13:654-79, Oct. 1934; Elec. Com. 13:159-71, Oct. 1934; Bell Mon. B-811.

310. Schelkunoff, S. A. Electromagnetic theory of coaxial transmission lines and cylindrical shields. B.S.T.J. 13:532-79, Oct. 1934; Bibliog. 22+ refs., pp. 78-9; Bell Mon. B-816.

311. ----------. Coaxial communication transmission lines. E.E. 53:1592-3, Dec. 1934; Bell Mon. B-831.

312. Hansell, C. W. Resonant lines for frequency control. E.E. 54:852-57, Aug. 1935.

313. ---------- and P. S. Carter. Frequency control by low power factor line circuits. I.R.E.Proc. 24:597-619, Apr. 1936.

314. Carter, P. S. and G. S. Wickizer. Ultra-high frequency transmission between RCA building and Empire State building in New York City. I.R.E.Proc. 24:1082-94, Aug. 1936.

315. Mason, W. P. and R. A. Sykes. Use of coaxial and balanced transmission lines in filters and wide-band transformers for high radio frequencies. B.S.T.J. 16:275-302, July 1937.

316. Reukema, L. E. Transmission lines at very high radio frequencies. E.E. 56:1002-1011, Aug. 1937.

317. Salzberg, B. On the optimum length for transmission lines as circuit elements. I.R.E.Proc. 25:1561-4, Dec. 1937.

318. Schelkunoff, S. A. Impedance concept and its application to problems of reflection, refraction, shielding, and power absorption. B.S.T.J. 17:17-48, Jan. 1938.

319. Wentz, J. F. Transmission characteristics of the coaxial structure. Bell Labs. Rec. 16:196-200, Feb. 1938.

320. Strieby, M. E. Coaxial cable system for television transmission. B.S.T.J. 17:438-57, July 1938.

321. Burrows, C. R. Exponential transmission line. B.S.T.J. 17:555-73, Oct. 1938.

322. Wheeler, H. A. Transmission lines with exponential taper. I.R.E.Proc. 27:65-71, Jan. 1939.

323. Nergaard, L. S. and B. Salzberg. Resonance impedance of transmission lines. I.R.E.Proc. 27:579-84, Sept. 1939.

324. Strieby, M. E. and J. F. Wentz. Television transmission over wire lines. B.S.T.J. 20:62-81, Jan. 1941.

325. ---------- and C. L. Weis. Television transmission. I.R.E.Proc. 29:371-81, July 1941.

326. Smith, P. H. L-type impedance transforming circuits. Electronics 15:48-54,125, 3 Mar. 1942.

U H F RADIATION

Periodical articles:

327. Carson, J. R. Generalization of the reciprocal theorem. B.S.T.J. 3:393-9, July 1924; Bell Mon. B-62-1.
328. ----------. Reciprocal theorems in radio communication. I.R.E.Proc. 17:952-6, June 1929; Bell Mon. B-400.
329. ----------. Reciprocal energy theorem. B.S.T.J. 9:325-31, Apr. 1930; Bell Mon. B-492.
330. Fry, T. C. Plane waves of light. I. Electromagnetic behavior. J.O.S.A. 15:137-61, Sept. 1927; Bell Mon. B-272; II. Reflection and refraction. J.O.S.A. 16:1-25, Jan. 1928; Bell Mon. B-286; III. Absorption by metals. J.O.S.A. 22:307-22, June 1932; Bell Mon. B-676.
331. Carter, P. S. Circuit relations in radiating systems and applications to antenna problems. I.R.E.Proc. 20:1004-1041, June 1932.
332. Schelkunoff, S. A. Some equivalence theorems of electromagnetics and their application to radiation problems. B.S.T.J. 15:92-112, Jan. 1936; Bell Mon. B-914.
333. Stratton, J. A. and L. J. Chu. Diffraction theory of electromagnetic waves. Phys. Rev. 56:99-107, 1 July 1939.
334. Schelkunoff, S. A. General radiation formula. I.R.E. Proc. 27:660-66, Oct. 1939.
335. Stratton, J. A. and L. J. Chu. Steady-state solutions of electromagnetic field problems. I. Forced oscillations of a cylindrical conductor; II. Forced oscillations of a conducting sphere; III. Forced oscillations of a prolate spheroid. J.A.P. 12:230-48, Mar. 1941.
336. Kandoian, A. G. Radiating systems and wave propagation. Electronics 15:4:39-44, Apr. 1942; Bibliog. 27 refs., p. 44.

U H F RADIATION: Antennas and arrays

Books:

337. Keen, R. Wireless direction finding. Ed. 3, 1938.
338. Brückmann, H. Antennen: ihre theorie und technik, 1939.
339. Harper, A. E. Rhombic antenna design, 1941.

Periodical articles:

340. Foster, R. M. Directive diagrams of antenna arrays. B.S.T.J. 5:292-307, Apr. 1926; Bell Mon. B-194.
341. Yagi, H. Beam transmission by ultra-short wave. I.R.E. Proc. 16:715-41, 1928.
342. Pistolkors, A. A. Radiation resistance of beam antennas. I.R.E.Proc. 17:562-628, Mar. 1929.
343. Southworth, G. C. Certain factors affecting the gain of directive antennas. I.R.E.Proc. 18:1502-36, Sept. 1930; Bibliog. pp. 1532-6; Bell Mon. B-539; also in B.S.T.J. 10:63-95, Jan. 1931.
344. Sterba, E. J. Theoretical and practical aspects of directional transmitting systems. I.R.E.Proc. 19: 1184-1215, July 1931; Bell Mon. B-592.
345. Bruce, E. Developments in short-wave directive antennas. I.R.E.Proc. 19:1406-33, Aug. 1931; also in B.S.T.J. 10:656- , Oct. 1931.
346. Carter, P. S., C. W. Hansell and N. E. Lindenblad. Development of directive transmitting antennas by RCA Communications, Inc. I.R.E.Proc. 19:1773-1842, Oct. 1931.
347. Carter, P.S., Circuit relations in radiating systems and applications to antenna problems. I.R.E.Proc. 20: 1004-1041, June 1932.
348. Chinn, H. A. and J. A. Stratton. Radiation characteristics of a vertical half-wave antenna. I.R.E.Proc. 20:1892-1913, Dec. 1932.
349. Feldman, C. B. Optical behavior of the ground for short radio waves. I.R.E.Proc. 21:764-801, June 1933.
350. Friis, H. T., C. B. Feldman and H. M. Sharpless. Determination of the direction of arrival of short radio waves. I.R.E.Proc. 22:47-78, Jan. 1934; Bell Mon. B-772.
351. Kear, F. G. Maintaining the directivity of antenna arrays. I.R.E.Proc. 22:847-69, July 1934; (M.I.T.E.E. Dpt. Sc. D. Thesis).
352. Bruce, E., L. R. Beck and A. C. Lowry. Horizontal rhombic antennas. B.S.T.J. 14:135-58, Jan. 1935; Bell Mon. B-835.
353. Barrow, W. L. On the impedance of a vertical half-wave antenna above an earth of finite conductivity. I.R.E.Proc. 23:150-67, Feb. 1935.
354. Brown, G. H. Directional antennas. I.R.E.Proc. 25:78-145, Jan. 1937.
355. Friis, H. T. and C. B. Feldman. Multiple unit steerable antenna for short-wave reception. B.S.T.J. 16:337-419, July 1937.

356. Foster, Donald. Radiation from rhombic antennas. I.R.E. Proc. 25:1327-53, Oct. 1937.
357. Carter, P. S. Charts for transmission line measurements and computations. (For use with antenna tests.) R.C.A. Rev. 3:355-68, Jan. 1939.
358. Lindenblad, N. E. Television transmitting antenna for Empire State Building. RCA Rev. 3:387-408, Apr. 1939.
359. Carter, P. S. Simple television antennas. RCA Rev. 3:168-85, Oct. 1939.
360. Kraus, J. D. Antenna arrays with closely spaced elements. I.R.E.Proc. 28:76-84, Feb. 1940.
361. ----------. Corner-reflector antenna. I.R.E.Proc. 28:513-19, Nov. 1940.
362. Brown, G. H. and J. Epstein. UHF antenna of simple construction. Communications 20:7:3-5, July 1940.
363. Lindenblad, N. E. Antennas and transmission lines at Empire State television station. I. Communications 20:5:13- , May 1940; II. Communications 21:4:10-14,24-6, Apr. 1941; III. Communications 21:5:9-13,31-3, May 1941.
364. Brown, G. H. Turnstile antenna for ultra-high frequency broadcasting. (Abstract). I.R.E.Proc. 29:221, Apr. 1941.
365. Schelkunoff, S. A. Theory of antennas of arbitrary size and shape. I.R.E.Proc. 29:493-521, Sept. 1941.

U H F RADIATION: Horns

Periodical articles:

366. Barrow, W. L. and L. J. Chu. Theory of the electromagnetic horn. I.R.E.Proc. 27:51-64, Jan. 1939.
367. Barrow, W. L. and F. D. Lewis. Sectoral electromagnetic horn. I.R.E.Proc. 27:41-50, Jan. 1939.
368. Southworth, G. C. and A. P. King. Metal horns as directive receivers of ultra-short waves. I.R.E.Proc. 27:95-102, Feb. 1939.
369. Reber, G. Electromagnetic horns. Communications 19:2:13-15, Feb. 1939.
370. Barrow, W. L. and L. J. Chu. Electromagnetic horn design. E.E. 59:333-8, July 1939.
371. Barrow, W. L., L. J. Chu and J. J. Jansen. Biconical electromagnetic horns. I.R.E.Proc. 27:769-79, Dec. 1939.
372. Barrow, W. L. and C. Shulman. Multiunit electromagnetic horns. I.R.E.Proc. 28:130-36, Mar. 1940.
373. King, A. P. Metal horns as radiators of electric waves. Bell Labs. Rec. 18:247-50, Apr. 1940.

U H F WAVE GUIDES

Periodical articles:

374. Rayleigh, Lord. On the passage of electric waves through tubes, or the vibrations of dielectric cylinders. Phil. Mag. 43:25-32, 1897.
375. Hondros, D. and P. Debye. Elektromagnetische Wellen an Dielektrischen Drähten. Ann.d.Phys. 32:465-76, 1910.
376. Zahn, H. Über den Nachweis Elektronmagnetischer Wellen an Dielektrischen Drähten. Ann.d.Phys. 49:907-33, 1916.
377. Schriever, O. Elektromagnetische Wellen an Dielektrischen Drähten. Ann.d.Phys. 63:645-73, 1920.
378. Carson, J. R. Guided and radiated energy in wire transmission. A.I.E.E.J. 43:908-13, Oct. 1924.
379. Bergmann, L. and L. Kruegel. Messungen in Strahlungsfeld einer in Innern eines metallischen Hohlzylinders erregten Linear-antennen. Ann.d.Phys. 21:113-38, Oct. 1934.
380. Southworth, G. C. Hyperfrequency wave guides - general considerations and experimental results. B.S.T.J. 15:284-309, Apr. 1936; Bell Mon. B-927.
381. Carson, J. R., S. P. Mead and S. A. Schelkunoff. Hyperfrequency wave guides - mathematical theory. B.S.T.J. 15:310-33, Apr. 1936; Bell Mon. B-921.
382. Brillouin, L. Propagation d'ondes electromagnetiques dans un tuyua. R.G.E. 40:227-39, 22 Aug. 1936.
383. Barrow, W. L. Transmission of electromagnetic waves through hollow tubes of metal. I.R.E.Proc. 24:1298-1328, Oct. 1936.
384. Southworth, G. C. Some fundamental experiments with wave guides. I.R.E.Proc. 25:807-22, July 1937.
385. ----------. New experimental methods applicable to ultra-short waves. J.A.P. 8:660-65, Oct. 1937.
386. Page, L. and N. I. Adams, Jr. Electromagnetic waves in conducting tubes. Phys. Rev. 52:647-51, Sept. 1937.
387. Schelkunoff, S. A. Transmission theory of plane electromagnetic waves. I.R.E.Proc. 25:1457-92, Nov. 1937; Bell Mon. B-1029.
388. Brillouin, L. Theoretical study of dielectric cables. Elec. Com. 16:350-72, Apr. 1938.
389. Schelkunoff, S. A. A note on certain guided waves in slightly noncircular tubes. J.A.P. 9:484-8, July 1938.
390. Chu, L. J. Electromagnetic waves in elliptic hollow pipes of metal. J.A.P. 9:583-91, Sept. 1938.
391. Brillouin, L. Propagation des ondes sur les cables diélectriques. Soc.Fr.d'Elec.Bul. :899-932, Oct. 1938.

392. Chu, L. J. and W. L. Barrow. Electromagnetic waves in hollow metal tubes of rectangular cross section. I.R.E.Proc. 26:1520-55, Dec. 1938.

393. Barrow, W. L. and F. M. Greene. Rectangular hollow-pipe radiators. I.R.E.Proc. 26:1498-1519, Dec. 1938.

394. Clavier, G. Theoretical relationships of dielectric guides (cylindrical) and coaxial cables. Elec. Com. 17:276-90, Jan. 1939. This article was based on two earlier articles: Theories des cables diélectriques cylindriques et relations avec la théorie des cables coaxiaux. Soc.Fr.d'Elec.Bul. (5) 8:355-88, Apr. 1938. Etude complèmentaire sur les coefficients d'attenuation dans les cables diélectriques cylindriques et les cables coaxiaux. Soc.Fr.d'Elec. Bul. (5) 8:793-804, Sept. 1938.

395. ---------- and V. Altovsky. Recherches expérimentales sur la propagation des ondes électromagnétiques dans les guides cylindriques. R.G.E. 55:697-706, 27 May, 1939; R.G.E. 55:731-43, 3 June 1939.

396. ----------. Translation of above. Experimental researches on propagation of electromagnetic waves in dielectric (cylindrical) guides. Elec. Com. 18:81-105, July 1939.

397. Lamont, H. R. L. Theory of resonance in microwave transmission lines with discontinuous dielectric. Phil. Mag. 29:521-40, June 1940; Bibliog. 21 refs., pp. 539-40.

398. Chu, L. J. Calculation of the radiation properties of hollow pipes and horns. J.A.P. 11:603-10, Sept. 1940.

399. Kemp, J. Electromagnetic waves in metal tubes of rectangular cross section. I.E.E.J. 88:III:213-18, Sept. 1941; also in Elec. Com. 20:2:73-9, 1941.

U H F WAVE GUIDES: Resonators

Periodical articles:

400. Reber, G. Electric resonance chambers. Communications 18:5-8, Dec. 1938.

401. Hansen, W. W. A type of electrical resonator. J.A.P. 9:654-63, Oct. 1938.

402. ----------. On the resonant frequency of closed concentric lines. J.A.P. 10:38-46, Jan. 1939.

403. Borgnis, F. Elektromagnetische Eigenschwingungen dielektrischer Raume. Ann.d.Phys. (5) 35:359-84, 1939.

404. Jouquet, M. Sur les oscillations électromagnetiques naturelles d'une cavité. Comptes Rendus 209:25-7, 203-4, 1939.

405. Barrow, W. L. and W. W. Mieher. Natural oscillations of electrical cavity resonators. I.R.E.Proc. 28:184-90, Apr. 1940.

406. Hahn, W. C. New Method for calculation of cavity resonators. J.A.P. 12:62-8, Jan. 1941.

407. Condon, E. U. Forced oscillations in cavity resonators. J.A.P. 12:129-32, Feb. 1941.

U H F PROPAGATION

Book:

408. Pederson, P. O. Propagation of radio waves along the surface of the earth and in the atmosphere. Copenhagen, 1927.

Periodical articles:

409. Breit, G. and M. A. Tuve. Test of the existence of the conducting layers. Phys. Rev. 28:554-75, Sept. 1926.

410. Appleton, E. V. On some measurements of the equivalent height of the atmospheric ionised layer. Proc. Roy. Soc. (A) 126:542-69, Mar. 3, 1930.

411. Stratton, J. A. Effect of rain and fog on the propagation of very short radio waves. I.R.E.Proc. 18:1064-74, June 1930.

412. Jouast, R. Some details relating to propagation of very short waves. I.R.E.Proc. 19:479088, Mar. 1931.

413. Beverage, H. H., H. P. Peterson, and C. W. Hansell. Application of frequencies above 30,000 kilocycles to communication problems. I.R.E.Proc. 19:1313-33, Aug. 1931.

414. Trevor, B. and P. S. Carter. Notes on propagation of waves below ten meters in length. I.R.E.Proc. 21:387-426, Mar. 1933.

415. Schellung, J. C., C. R. Burrows, and E. B. Ferrell. Ultra-short-wave propagation. I.R.E.Proc. 21:427-63, Mar. 1933.

416. Englund, C. R., A. B. Crawford, and W. W. Mumford. Some results of study of ultra-short-wave transmission phenomena. I.R.E.Proc. 21:464-92, Mar. 1933.

417. Feldman, C. B. Optical behavior of the ground for short radio waves. I.R.E.Proc. 21:764-801, June 1933.

418. Friis, H. T., C. B. Feldman, and W. M. Sharpless. Determination of direction of arrival of short radio waves. I.R.E.Proc. 22:47-78, Jan. 1934.

419. Hull, R. A. Air mass conditions and bending of ultra-high-frequency waves. Q S T 19:13-18, June 1935.

420. Englund, C. R., A. B. Crawford, and W. W. Mumford. Further studies of ultra-short-wave transmission phenomena. B.S.T.J. 14:369-87, July 1935.
421. Burrows, C. R., A. Decino, and L. E. Hunt. Ultra-short-wave propagation over land. I.R.E.Proc. 23:1507-35, Dec. 1935.
422. Wwedensky, B. Diffractive propagation of radio waves.
I. Tech. Phys. USSR v. 2, pp. 624-39, 1935;
II. Tech. Phys. USSR v. 3, pp. 915-25, 1936;
III. Tech. Phys. USSR v. 4, pp. 579-91, 1937.
423. Norton, K. A. Propagation of radio waves over the surface of the earth and in the upper atmosphere. I. Ground wave propagation from short antennas. I.R.E.Proc. 24:1367-87, Oct. 1936; II. Propagation from vertical, horizontal, and loop antennas over a plane earth of finite conductivity. I.R.E.Proc. 25:12-3-36, Sept. 1937.
424. Clavier, A. G. Propagation tests with micro-rays. Elec. Com. 15:211-19, Jan. 1937.
425. Colwell, R. C. and A. W. Friend. The lower ionosphere. Phys. Rev. 50:632-5, Oct. 1, 1936.
426. Beverage, H. H. Some notes on UHF propagation. RCA Rev. 1:76-87, Jan. 1937.
427. Burrows, C. R. Radio propagation over plane earth-field strength curves. B.S.T.J. 16:45-7, Jan. 1937.
428. Mimno, H. R. Physics of the ionosphere. Rev. Mod. Phys. 9:1-43, Jan. 1937.
429. Eckersley, T. L. Ultra-short-wave refraction and diffraction. I.E.E.J. 80:286-304, Mar. 1937.
430. Smith, N. Extension of normal-incidence ionosphere measurements to oblique-incidence radio transmission. N.B.S.J. Res. 19:89-94, July 1937; Res. Paper, 1013.
431. Watt, R. A. W., A. F. Wilkins, and E. G. Bowen. Return of radio waves from the middle atmosphere. I. Proc. Roy. Soc. (A) 161:181-96, July 15, 1937.
432. Norton, K. A. Physical reality of space and surface waves in the radiation field of radio antennas. I.R.E. Proc. 25:1192-1202, Sept. 1937; Bibliog. 17 refs., pp. 1201-2.
433. Friend, A. W. and R. C. Colwell. Measuring the reflecting regions of the troposphere. I.R.E.Proc. 25:1531-41, Dec. 1937.
434. Burrows, C. R., A. Decino, and L. E. Hunt. Stability of 2-meter waves. I.R.E.Proc. 26:516-28, May 1938.
435. Englund, C. R., A. B. Crawford, and W. W. Mumford. Ultra-short-wave transmission and atmospheric irregularities. B.S.T.J. 17:489-519, Oct. 1938.

436. George, R. W. Study of ultra-high-frequency wide-band propagation characteristics. I.R.E.Proc. 27:28-35, Jan. 1939; also in RCA: Radio at UHF, 1940, pp. 90-106.

437. Katzin, M. Ultra-high-frequency propagation. Proc. Radic Club of Amer., Sept. 1939; also in RCA: Radio at UHF, 1940, pp. 107-42; Bibliog. 38 refs., pp. 140-42.

438. Waynick, A. H. Experiments on the propagation of ultra-short radio waves. I.R.E.Proc. 28:468-75, Oct. 1940.

439. Norton, K. A. Calculation of ground-wave field intensity over a finitely conducting spherical earth. I.R.E.Proc. 29:623-39, Dec. 1941.

SUBJECT INDEX

FOR BIBLIOGRAPHY

(Numbers refer to Reference Number)

INDEX

(index continued on following page)