ELECTRONIC INSTRUMENTS

Edited by

IVAN A. GREENWOOD, JR.
RESEARCH GROUP LEADER
GENERAL PRECISION LABORATORY, INC.

J. VANCE HOLDAM, JR.
DIRECTOR OF RESEARCH
LABORATORY FOR ELECTRONICS, INC.

DUNCAN MACRAE, JR.
TEACHING FELLOW, HARVARD UNIVERSITY

OFFICE OF SCIENTIFIC RESEARCH AND DEVELOPMENT
NATIONAL DEFENSE RESEARCH COMMITTEE

FIRST EDITION

RADIATION · LABORATORY · M·I·T · CAMBRIDGE · MASS · 1940-1945

NEW YORK · TORONTO · LONDON
McGRAW-HILL BOOK COMPANY, INC.
1948

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RADIATION LABORATORY SERIES

---

1. Radar System Engineering—*Ridenour*
2. Radar Aids to Navigation—*Hall*
3. Radar Beacons—*Roberts*
4. Loran—*Pierce, McKenzie, and Woodward*
5. Pulse Generators—*Glasoe and Lebacqz*
6. Microwave Magnetrons—*Collins*
7. Klystrons and Microwave Triodes—*Hamilton, Knipp, and Kuper*
8. Principles of Microwave Circuits—*Montgomery, Dicke, and Purcell*
9. Microwave Transmission Circuits—*Ragan*
10. Waveguide Handbook—*Marcuvitz*
11. Technique of Microwave Measurements—*Montgomery*
12. Microwave Antenna Theory and Design—*Silver*
13. Propagation of Short Radio Waves—*Kerr*
14. Microwave Duplexers—*Smullin and Montgomery*
15. Crystal Rectifiers—*Torrey and Whitmer*
16. Microwave Mixers—*Pound*
17. Components Handbook—*Blackburn*
18. Vacuum Tube Amplifiers—*Valley and Wallman*
19. Waveforms—*Chance, Hughes, MacNichol, Sayre, and Williams*
20. Electronic Time Measurements—*Chance, Hulsizer, MacNichol, and Williams*
21. Electronic Instruments—*Greenwood, Holdam, and MacRae*
22. Cathode Ray Tube Displays—*Soller, Starr, and Valley*
23. Microwave Receivers—*Van Voorhis*
24. Threshold Signals—*Lawson and Uhlenbeck*
25. Theory of Servomechanisms—*James, Nichols, and Phillips*
26. Radar Scanners and Radomes—*Cady, Karelitz, and Turner*
27. Computing Mechanisms and Linkages—*Svoboda*
28. Index—*Henney*

ELECTRONIC INSTRUMENTS

Edited by

IVAN A. GREENWOOD, JR.
RESEARCH GROUP LEADER
GENERAL PRECISION LABORATORY, INC.

J. VANCE HOLDAM, JR.
DIRECTOR OF RESEARCH
LABORATORY FOR ELECTRONICS, INC.

DUNCAN MACRAE, JR.
TEACHING FELLOW, HARVARD UNIVERSITY

OFFICE OF SCIENTIFIC RESEARCH AND DEVELOPMENT
NATIONAL DEFENSE RESEARCH COMMITTEE

FIRST EDITION

RADIATION · LABORATORY · M·I·T · CAMBRIDGE · MASS · 1940·1945

NEW YORK · TORONTO · LONDON
McGRAW-HILL BOOK COMPANY, INC.
1948

ELECTRONIC INSTRUMENTS

PRINTED IN THE UNITED STATES OF AMERICA

THE MAPLE PRESS COMPANY, YORK, PA.

Foreword

THE tremendous research and development effort that went into the development of radar and related techniques during World War II resulted not only in hundreds of radar sets for military (and some for possible peacetime) use but also in a great body of information and new techniques in the electronics and high-frequency fields. Because this basic material may be of great value to science and engineering, it seemed most important to publish it as soon as security permitted.

The Radiation Laboratory of MIT, which operated under the supervision of the National Defense Research Committee, undertook the great task of preparing these volumes. The work described herein, however, is the collective result of work done at many laboratories, Army, Navy, university, and industrial, both in this country and in England, Canada, and other Dominions.

The Radiation Laboratory, once its proposals were approved and finances provided by the Office of Scientific Research and Development, chose Louis N. Ridenour as Editor-in-Chief to lead and direct the entire project. An editorial staff was then selected of those best qualified for this type of task. Finally the authors for the various volumes or chapters or sections were chosen from among those experts who were intimately familiar with the various fields, and who were able and willing to write the summaries of them. This entire staff agreed to remain at work at MIT for six months or more after the work of the Radiation Laboratory was complete. These volumes stand as a monument to this group.

These volumes serve as a memorial to the unnamed hundreds and thousands of other scientists, engineers, and others who actually carried on the research, development, and engineering work the results of which are herein described. There were so many involved in this work and they worked so closely together even though often in widely separated laboratories that it is impossible to name or even to know those who contributed to a particular idea or development. Only certain ones who wrote reports or articles have even been mentioned. But to all those who contributed in any way to this great cooperative development enterprise, both in this country and in England, these volumes are dedicated.

L. A. DUBRIDGE.

ELECTRONIC INSTRUMENTS

EDITORIAL STAFF

Britton Chance	Editor
Ivan A. Greenwood, Jr.	Editor
Duncan MacRae, Jr.	Volume Editor
J. Vance Holdam, Jr.	Volume Editor

CONTRIBUTING AUTHORS

Britton Chance
Sidney B. Cohen
Arden H. Fredrick
Warren F. Goodell, Jr.
John W. Gray
William D. Green, Jr.
Ivan A. Greenwood, Jr.
J. Vance Holdam, Jr.
Arthur C. Hughes, Jr.
Andrew B. Jacobsen
George L. Kreezer
Richard W. Lee
John J. Lentz
Edward F. MacNichol, Jr.
Duncan MacRae, Jr.
Warren G. Proctor
Harry J. Reed
J. Robert Rogers
Wilfred Roth

Preface

THE title of the present volume, *Electronic Instruments*, carries with it the implied adjective "some." The specific kinds of electronic instruments which are treated are *electronic analogue computers, instrument servomechanisms, voltage and current regulators*, and *pulse test equipment.* Aside from the common denominator indicated by the title, these four types of equipment share claims as important adjuncts to many modern radar systems.

It has been the object of the authors to present both theoretical background and practical details of these instruments as they were known to the radar engineer. However, the authors believe firmly that radar applications of these devices represent only a very small part of their field of utility. An attempt has been made to emphasize the sort of information which the authors felt would have been most helpful to them when they were required to solve problems in the fields represented by this volume.

The preservation of the material of this volume was made possible through the foresight of I. I. Rabi and L. A. DuBridge, who appointed a committee consisting of L. J. Haworth, G. E. Valley, and B. Chance to consider the scope and content of a series of books on the general subject of electronic circuits. Volumes 17 to 22 of the present series are the result of the committee's survey. At the termination of hostilities an intensive writing program, under the leadership of L.N. Ridenour, was put into effect. Because of the rapid dissolution of the Radiation Laboratory, an accelerated writing schedule, using as many authors as possible, was unavoidable. Even with such a policy many authors made real sacrifices in giving up or postponing positions and fellowships in order to complete their contributions. The combination of accelerated schedules and divided efforts has regrettably resulted in discontinuities in the scope and treatment of the material covered, and in many cases has led to historical study inadequate for the proper assignment of credit for developments reported.

At the termination of the Radiation Laboratory Office of Publications, some of the writing, much of the editing, and all of the proofreading of the present volume still remained to be done. Credit for the

v

completion of this volume belongs in a large degree to the General Precision Laboratory Inc., of Pleasantville, N.Y. This Laboratory generously permitted one of the technical editors of the present volume to devote a large fraction of his time over a period of many months to the project, and made available extensive secretarial and drafting assistance. In addition, much of the manuscript was read and criticized by other members of the staff. Credit is due the many authors who assisted in the checking of proofs long after they had left the employ of the Radiation Laboratory.

Many of the developments described in this volume are contributions from laboratories in the United Kingdom. It is a pleasure to acknowledge the unstinting support of these British laboratories, and especially of Telecommunications Research Establishment (TRE). Through their generosity, several experts have visited this country and have contributed much useful information to this and other volumes of the Radiation Laboratory Series. Our gratitude for this international cooperation is due Sir Robert Watson Watt, W. B. Lewis, B. V. Bowden, F. S. Barton, F. C. Williams, and N. F. Moody, and their associates.

Background material on which parts of this volume are based was contributed by H. S. Sack of Cornell University.

The preparation of manuscript and drawings would have been impossible without the help of the production department under C. Newton; the Technical Coordination Group, under Dr. Leon Linford; the typing pool, under M. Dolbeare and P. Phillips; and the drafting room, under Dr. V. Josephson. The authors wish to acknowledge the invaluable help of the following editorial assistants, production assistants, and secretaries: Louise Rosser, Nora Van Der Groen, Joan Brown, Helene Benvie, Teresa Sheehan, Joan Leamy, Barbara Davidson, and Helen Siderwicz, all of the Radiation Laboratory, and Mary Pollock, Nora Applegate, and Gordon Clift of General Precision Laboratory.

THE AUTHORS.

CAMBRIDGE, MASS.,
October, 1946.

Contents

PART III. VOLTAGE AND CURRENT REGULATORS

PART IV. PULSE TEST EQUIPMENT

PART V

PART I

ELECTRONIC ANALOGUE COMPUTERS

CHAPTER 1

INTRODUCTION

By I. A. Greenwood, Jr.

1·1. General Comments on Computers.—Important parts of the technology that man is developing are the understanding of relationships as formalized in the science of mathematics and the application of the techniques of this science. In recent years considerable progress has been made in the development of devices that aid directly in the application of mathematical techniques. One class of such devices is known as "computers," also referred to as "calculating machines," "calculators," etc.

A computer may be defined as a device that performs mathematical operations on input data to yield new and generally more useful results. The abacus, for example, is a simple kind of manually operated computer that has been used for centuries by the Chinese and others as an adding machine. The input data of this device are the separate numbers that are entered by sliding beads according to definite rules. If this manipulation is done properly, the mathematical operation of addition of numbers is performed, yielding the total in the form of numbers represented by bead positions. Computers today range in complexity from devices as simple as the abacus to machines filling large rooms with many thousands of intricate parts and capable of solving rapidly problems of a very high order of complexity or capable of solving simpler problems in large quantities with considerable economic savings as compared with other less mechanized methods of calculations.

Applications of Computers.—The Radiation Laboratory has been concerned with computers because of the important and intimate relationship of computers and military radars. Computers have been used as integral parts of systems built around radar for such applications as blind bombing, navigation, control of gunfire, radar trainers, etc.[1] Pos-

[1] The unique characteristics of radar have usually made it necessary or at least desirable to use special computers designed to use the radar information to fullest advantage. For example, early ship-to-ship fire-control computers were designed to take full advantage of accurate optical azimuth data but to accept poor range data; fire-control radar, with its excellent range accuracy but only fair azimuth accuracy required completely different computers for maximum effectiveness. Radars are also modified for integration with computers. For example, the LAB (low-altitude bombing) radar and computer equipment that achieved such spectacular results against

sible uses for computers are both numerous and of far-reaching technical consequence. Vannevar Bush has remarked[1] that "the world has arrived at an age of cheap complex gadgets of great reliability," citing 30-cent radio tubes as examples. This quotation, the authors believe, may well be used to summarize the present or near-future status of the computer field. Of course, "cheap" is here understood to mean cheap in comparison with savings effected or in relation to value of goods produced with the assistance of computers. Among the possible uses of computers are numerical solution of scientific and engineering problems,[2] industrial process control computations, conduct of purely mathematical research, transformations of data in physical measurements, and, in general, substitution of mechanization for specialized human operations. The process control field appears to be particularly fertile and is relatively unexploited from the standpoint of equipment whose operation is based on more than just simple linear functions of a very limited number of measured variables. The design of large aircraft is providing an expanding field for computer applications, for lightweight devices improving the efficiency and safety of aircraft operation are of great economic importance. It is anticipated that computers will be developed and used in large aircraft for automatic solutions of all or major parts of the following problems: navigation, air-traffic control, engine efficiency, etc.

A word regarding the more distant future may be of interest, even if risky. Without exceeding a reasonable extrapolation of known techniques, one may speculate on the possibilities of desk-size machines containing the equivalent of whole libraries and capable of high-speed selection and cross indexing,[3] machines that perform simple associative reasoning, machines that type spoken words, machines that translate one language into another, etc. Even that favorite of the cartoonists, the "mechanical man" that can beat its human master in a chess game, cannot be said to be an impossibility.

Reasons for Using Computers.—Computers are presently used and can be used for a variety of reasons. In some cases the problems that must be solved are too difficult for simple methods of numerical solution. A differential analyzer,[4] for example, is capable of giving numerical solu-

Japanese shipping in the recent war used a special bombardier's radar indicator with computer-controlled sweeps and electronic markers.

[1] V. Bush, "As We May Think," *Atlantic Monthly*, **176**, No. 7, 101–108, July 1945.

[2] An example chosen at random from many such devices described in the literature is C. E. Berry, *et al.*, "A Computer for Solving Linear Simultaneous Equations," *Jour. Applied Phys.*, **17**, 262–272, April 1946.

[3] The "memex" of V. Bush's stimulating article, *loc. cit.*

[4] See, for instance, V. Bush and S. H. Caldwell, "A New Type of Differential Analyser." *Jour. Franklin Inst.*, **240**, 255, October 1945.

tions of differential equation problems that are impractical (usually for economic reasons) to solve by other means. This valuable research tool has stimulated activity in the mathematical fields where ordinary processes of analysis are severely limited. As Bush and Caldwell point out, the way is open for the creation and practical use of new functions defined only by their differential equations. To this should be added functions defined only by implicit equations.[1] Certain mathematical processes, such as adding long columns of figures, may be simple but very tedious; devices such as adding machines are useful for handling this type of work. Sometimes a very rapid solution of an involved problem is required. For example, compilation of a ballistic table by unmechanized computing methods might take man-years to complete; with the right computing equipment the entire job can be done with a few man-days or man-weeks. Computers are often used merely to simplify complicated procedures or to present data in a more convenient form. One important advantage of many computers—and particularly so of those which this book will emphasize—is that they will do their work unattended or with little manual supervision. For example, the computing mechanisms associated with radar or visual bombsights will do a large part of the "thinking" that is needed to solve the somewhat involved geometry of the bombing problem. The ability of computers to work unattended is of utmost importance in the field of automatic process control, where it may be necessary to perform certain mathematical operations on input data in order to arrive at the correct information for actuating the control member of the system. The ability of computers to work unattended is a major economic consideration in this field.

1·2. Types of Computers and Computer Elements.—It is of interest to classify the various types of computers and computer elements. Such a classification will be referred to in the following section where the scope of this treatment will be discussed. Four useful types of classification are

1. Automatic vs. manually operated.
2. Electronic and electromechanical vs. purely mechanical.
3. Digital vs. analogue.
4. Single-purpose vs. multipurpose.

Manual vs. Automatic Computers.—The abacus has already been listed as a simple example of a manually operated computer, without definition. A "manually operated computer" is defined as one in which most of the

[1] It is true that such equations can usually be written as differential equations and solved as such on a differential analyzer. This particular type of mathematical rearrangement, however, is not essential to the design of computers solving implicit equations by the techniques to be described; hence the distinction.

manipulation of data is done by manual operation, while an "automatic computer" is one in which most of the manipulation of data is done without manual assistance other than entering the input data and receiving the output data. Thus, a simple slide rule is a manually operated computer. If a series of slide rules were connected together by motors and other gadgetry to perform something more than the simple operations that may be done with a slide rule by itself, then there would result an example of an automatic computer. The distinction between the two computer types, although admittedly not clear-cut, does serve a useful purpose in roughly subdividing the general field of computing mechanisms. As will be explained below, the contents of Part I of this volume are limited principally to automatic computers and computer elements.

Electronic and Electromechanical vs. Purely Mechanical Computer Elements.—This distinction is of interest, for it means a major difference in the background and facilities required to design and produce these two types of computer elements. There are fairly distinct characteristics that can be associated with each type of element; these will be discussed in some detail in Sec. 2·11. The difference between these two types of computer elements is also important in connection with the emphasis of this treatment, as discussed in Sec. 1·3.

Digital vs. Analogue Computers.—In considering the subject of computers, one is immediately concerned with the concept of quantity and magnitude. There are three ideas associated with this concept: (1) the thing described, that is, length, voltage, etc.; (2) the unit, that is, feet, volts, etc.; and (3) the number of units, that is, 10 ft., 12 volts, etc. Quantity in a strict mathematical sense is a very abstract concept. In a computer the representation of quantity must be specific. In the way that it is specific, computers may be divided into two categories: those which deal with continuously variable physical magnitudes and those which deal with magnitudes expressed as a number of digits. The first type will be referred to as "analogue computers"; the second type will be referred to as "digital computers." Thus, the simple computer shown in Fig. 1·1 is an example of an analogue computer. The standard commercial desk calculating machine may be classified as a digital computer. Some of the newer high-speed electronic and electromechanical computers operate on the digital principle and will be discussed briefly below and in following chapters.

A fundamental distinction between digital and analogue computers is the fact that in digital computers the accuracy is limited *only by the number of significant figures provided for* whereas in analogue computers accuracy is limited by the *percentage errors of the devices used multiplied by the full ranges of the variables that they represent.* This difference is so fundamental and so important that it can scarcely be emphasized enough. A decision between the two radically different philosophies of design is the first step that is made in selecting or designing a computer for a specific purpose. If four or five or more significant figures are required in computations, analogue computers are usually simply not good enough and digital devices must be used. Where two, three, or four significant figures are all that are required, analogue computers may be far simpler than digital computers, and it would be foolish to pay in complexity for the digital computers' unused ability to handle more significant figures.

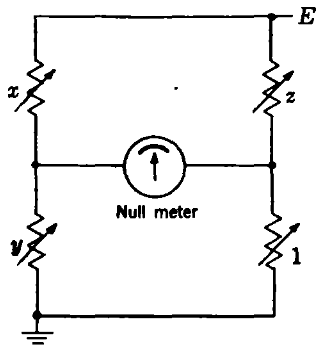


FIG. 1·1.—Simple bridge computer for $z = x/y$.

Analogue computers represent a general method that has numerous applications, namely, the use of one physical system as a model for another system, more difficult to construct or measure, that obeys equations of the same form. Examples of applications of the general method that might be called analogue computers but are generally not thought of as such are the use of electrolytic troughs to represent certain systems involving functions of a complex variable and the use of equivalent electrical networks to represent complicated differential equations.[1]

Single-purpose vs. Multipurpose Elements.—The type of computer generally used in bombsights, fire-control equipment, process control, etc., consists of a group of elements each of which performs a single mathematical operation, the data passing through each element only once. The presence of feedback loops is not considered an exception to this statement. Such elements may be called "single-purpose elements." Computers are also built on a different and very interesting philosophy. This type of computer involves a relatively small number of simple adding, subtracting, and memorizing elements. Data are switched through these elements in a complicated fashion by a control or programming center, which is usually operated by a coded tape or its equivalent. This concept has been used in many of the interesting high-speed digital computers that have been developed for the numerical solution of complex mathematical problems as aids to scientific and engineering researches. Among the important examples of this type of computer are the IBM-Harvard University electrically controlled mechanical com-

[1] G. Kron, "Numerical Solutions of Ordinary and Partial Differential Equations by Means of Equivalent Circuits," *Jour. Applied Phys.*, **16**, 172–186, March 1945.

puter[1] and the electronic computers developed at the University of Pennsylvania.[2]

When several sets of data are sent through a multipurpose element, it is necessary to keep them separated. Coding for identification of data may be accomplished

1. Through position in time of pulses relative to a timing pulse.
2. Through amplitude, width, or interval coding of pulses or by other wave-shape characteristics.
3. Through frequency differences.

Advantages of multipurpose computers at present are flexibility and, in the case of electronic types, extremely high speed. Electronic or electromechanical computers using multipurpose techniques are usually digital devices based on the binary system of numbers rather than the decimal system, since only the digits 0 and 1 are used in this system and can conveniently be represented by two states of a vacuum tube, namely, conducting and nonconducting, by the absence or presence of a pulse in a given time interval, or by the two positions of a relay armature. In pulse digital computers, numbers are thus represented in the binary system by a sequence of pulses and spaces, the pulses corresponding to 1's and the spaces corresponding to zeros. Numbers so represented may be "remembered" by injecting them into a system having a transmission delay greater than the period of the pulse sequence, the output of such a delay being amplified, resynchronized with a master timing standard, and reinjected to circulate in such closed cycles indefinitely until called up for use. A variety of other memory methods are used or under investigation.

The field of electronic pulse digital computers is one of great research interest at this time, for the combination of high-speed operation and ability to retain as many significant figures as desired offers exciting potentialities. It appears possible to build electronic pulse digital computers that operate on continuously varying data but in which the necessary computations are carried out repeatedly at closely spaced time intervals. This would give outputs equivalent for most purposes to those of continuous (analogue) computers, but with the advantages of speed and the extra significant figures that can be carried as compared with analogue devices. More is said about speed of computation in Sec. 1·4.

1·3. Limitations of Scope and Plan of Part I.—The treatment of this book is of necessity somewhat limited in scope. Limitations of time and personnel have made it seem desirable to concentrate the emphasis of this treatment on devices on which the Radiation Laboratory had worked or ones with which it was familiar. An attempt has been made, however,

[1] *A Manual of Operation for the Automatic Sequence Controlled Calculator*, Harvard University Press, 1946.

[2] See footnotes, Sec. 3·8, for references.

at least to mention other important devices, with references, so that the interested reader may pursue further the study of devices not discussed in detail.

The majority of the computers that Part I of this volume treats may be classified as *automatic, electronic or electromechanical, analogue computers, with single-purpose elements.* To the extent possible, information on mechanical elements has been included but is not complete or detailed. Excluded are such important aids to mathematical computation as the ordinary commercial desk calculating machine (a mechanical digital computer).

Following this introductory chapter will be found discussions on overall design procedure, important design principles, and a discussion of various representations of quantities and their characteristics. Other chapters will treat isolated operations, groups of operations, and complete computers.

Two goals have been set for Part I: a listing and explanation of a number of devices and methods and a systemization of the analogue computer field, including a scheme for "cataloguing" computer elements and methods, and a summary of the basic principles underlying their logical selection and combination.

1·4. Speed of Computations.—Von Neumann[1] points out that for a typical binary system multiplication problem carrying eight significant decimal digits some 1000 to 1500 steps are required. Using relays with reaction times of 5 msec (a lower limit for available relays) roughly 5 to 8 sec per eight decimal digit multiplication would be required. The same multiplication on a fast modern desk computing machine at present takes 10 sec, and for standard International Business Machines multipliers, the time is 6 sec. Some time may be saved in such a process at the expense of complexity of equipment. With vacuum tubes, a reaction time of 1 μsec now may be readily achieved. With this reaction time, the multiplication involving 1000 to 1500 steps would require only 1 to $1\frac{1}{2}$ msec, far less than any nonvacuum-tube device.[2] Though this discussion applies directly to digital computers, the details of which are outside the scope of this book, the general conclusion that electronic devices can be much faster than electromechanical or mechanical devices nevertheless applies to most computing problems. However, extreme speed is often not of great importance; and for a wide variety of computer purposes, the speeds available with electromechanical and mechanical devices are completely adequate. For the remaining cases where very

[1] J. Von Neumann, "First Draft of a Report on the EDVAC," Moore School, University of Pennsylvania. See also other references given in footnotes of Sec. 3·8.

[2] D. R. Hartree, "The ENIAC, An Electronic Calculating Machine," *Nature* (London), **157,** 527, Apr. 20, 1946, states that the multiplication of two numbers of 10 decimal digits takes "a few milliseconds" with these methods.

high speeds are required, the all-electronic devices are fortunately becoming available. In most of the computing devices with which Part I will be concerned, methods of computation are arranged so that only two to four significant figures are retained, and the isolated operations described have minimum reaction times ranging from a fraction of a millisecond to a few seconds.

1·5. The Computer Problem.—The computer design problem as it generally exists may be stated as follows: Given input data of a specified nature, range, speed, etc., design a device that will operate on these data to give output data of a specified nature, range, speed, etc., and that will operate at a specified accuracy under specified service conditions.

Usually one is faced with the problem of designing a specific computer for a specific task, and consideration of flexibility will give way to other considerations such as cost, weight, complexity, etc. In many computers just the opposite is true: examples are the MIT differential analyzer, the Harvard computer, the University of Pennsylvania vacuum-tube computers, and some commercial computing devices. Most of the devices with which the Radiation Laboratory has been concerned are classed as specialized computers, although in almost every case an attempt was made to use standard components or components suitable for standardization.

Throughout this book there will be frequent discussions of design factors applying to the design of some special device. In the interests of compactness, a master list of design factors has been prepared and will be found in Chap. 19. When practical, this list will be referred to for the majority of the design factors applying to any specific device, leaving for detailed discussion only those factors of special importance to the specific device. In applying the master design-factor list to any specific device, it will, of course, be found that many factors do not apply. These are generally obvious, however, and it is felt that the advantages of a reasonably complete design-factor "check list" will be of sufficient practical assistance to the designer to outweigh such lack of universality.

1·6. Summary.—Computers are defined; examples given; and uses discussed. Four types of classifications of computers and computer elements are mentioned: automatic vs. manual, electronic or electromechanical vs. mechanical, digital vs. analogue, and single-purpose vs. multipurpose. It is stated that the emphasis of Part I will be on automatic, electronic and electromechanical, analogue, single-purpose computers and computer elements. Speed of computing is mentioned, with the conclusion that where very high speeds are required, electronic computers should be considered but that for a very wide range of applications speeds available with electromechanical computers are adequate. The usual problem of computer design is mentioned.

CHAPTER 2

COMPUTER DESIGN

By I. A. Greenwood, Jr., and D. MacRae, Jr.[1]

2·1. Introduction.—The usual problem with which the computer designer is faced is discussed in Sec. 1·5. This chapter will summarize a systematic procedure for designing computers. Although this procedure is arbitrary and subject to the readers' revisions to suit the readers' needs and temperament, it has nevertheless been found to be both useful and expeditious in the long run. This procedure summary is followed by a presentation of some of the more important basic principles and techniques used in computers. The chapter concludes with a discussion of the subject of representation of quantity and includes a brief comparison of the characteristics of some of the more important types of data representation used in electronic and electromechanical computers.

DESIGN PROCEDURE

2·2. Summary of a Systematic Design Procedure.—Systematic procedure for designing electronic and electromechanical computers is presented in the present section. Reference is made to Chap. 7 wherein a sample computer design is discussed in terms of this design procedure. Reference is also made to Chap. 19, in which the subject of electronic engineering is discussed from a more general standpoint.

Preliminary Information.—In starting the design of a computer, the designer must have specific information as to what is to be computed. This statement sounds trite, yet it contains an important truth. The translation of an operational need, often vague, into computer specifications may be a difficult task, requiring a high caliber of professional judgment. It should preferably be done by or in close consultation with the person or team that is charged with the computer design. It frequently turns out that over-all specifications need revision after a design has been carried along. It is profitable for the designer always to keep in mind the operational need as well as the final over-all specifications. Included in the specifications should be such things as the characteristics of the inputs and outputs and a complete list of the design limitations

[1] Section 2·5 is by I. A. Greenwood, Jr., and D. MacRae, Jr. The rest of Chap. 2 by I. A. Greenwood, Jr.

and operating conditions that must be met (e.g., cost, design time, temperature, vibration, etc.). A comprehensive check list of design factors will be found in Chap. 19. In starting a design, a designer will want to be familiar with a variety of methods and devices for performing isolated mathematical operations and groups of operations and the basic principles and techniques for bringing together these methods and devices into a high-quality computer.

At this point, the designer will wish to draw an over-all functional block diagram and to formulate the fundamental computer equations. It is worth while to do this very carefully. Some of the methods of Sec. 2·3 may be found helpful in this process.

Creating a Block Diagram.—Having the over-all specification, the over-all functional block diagram, and the computer equations, the designer is next ready to create a detailed block diagram. The creation of a block diagram is an experimental technique in which the designer tries fitting together on paper the various blocks at his disposal, using the principles of implicit and explicit functions, error cancellation, etc., as discussed in Secs. 2·3 to 2·4, immediately following the present section. That combination of blocks which best satisfies the initial design requirements and applicable engineering considerations may be accepted as a starting block diagram. It will often be found that the best computer will combine a number of types of representations of data, such as mechanical displacements and rotations, voltages, and impedances.

Concurrently with the creation of a block diagram over-all scale factors should be roughly determined, as they may have a great deal to do with the choice of blocks. At this point also it is wise to investigate those components or circuits whose performance is critical to the success of the block diagram selected. An example of the application of this statement is found in Chap. 7.

In selecting blocks for a block diagram, it is important to consider both (1) the possible effects of the block being considered on other blocks and linking devices in the complete chain of operations of the computer and (conversely) (2) the possible effects of other blocks on the data inputs and outputs, scale factors, accuracy, etc., on the block being considered.

Preliminary Design.—With a block diagram chosen, a preliminary design may be made in some detail. The first step in this process is a detailed choice of system scale factors and assignment of permissible errors to the various blocks.

At this stage in the design it should be kept in mind that all components which are used in the design should be procurable or capable of being manufactured in the quantities desired. It is a common fault of designs to carry along too far a scheme using unprocurable components. This statement should not be interpreted as arguing against the use of special components where justified; however, where special components are used, they must be suitable for production in the quantities desired. Naturally, the smaller the quantities the more special may be the components. Very large quantities may also justify special components.

Detailed Performance Analysis.—A preliminary design having been completed, it should be analyzed in some detail for expected performance under all operating conditions wherever this is practical in order to make sure that all design considerations have been met satisfactorily and that reasonable compromises between conflicting factors have been effected.

Detailed Design.—If this analysis turns out satisfactorily, the design should be worked over in great detail in order to make certain that all elements used are consistent with the design considerations. Exact and complete specifications and tolerances for all component parts must be determined. The design is not completed until this is done; a single adverse characteristic of a component may require a new block diagram to be chosen. It is the practice of some successful laboratories to consider detailed production inspection procedures as essential parts of the complete specifications for all components and assembled equipments.

Construction of Model.—Either after the theoretical analysis or after the detailed analysis of the components, a breadboard or prototype model should be constructed. Two important functions of a model are (1) to serve as a check on design calculations and (2) to stimulate further thinking. Models also provide the means for obtaining information that is impractical to calculate. An example of such information is stray capacitance. A working model is always useful and almost always essential when production is being initiated.

Repetition of Steps.—It will generally be found that the entire process must be repeated perhaps several times before a really satisfactory final design is achieved.

Two of the most important factors in successful computer designing are (1) getting the feel for the proper type of data representation at each point in a chain of operations and (2) mastering the techniques of implicit solutions by means of feedback loops. In all computer design, experience and ingenuity assist one considerably. However, by the exercise of only a little extra patience in the process of fitting blocks together on paper in the creation of a block diagram and in the execution of detailed design, the reader new to the field should be able to design satisfactory computers for most technical purposes. The authors maintain that computer design can and should become a working tool of the modern engineer and scientist rather than a "black art" available only to a few initiates.

BASIC PRINCIPLES AND TECHNIQUES[1]

2·3. Rearrangement of Computer Equations.—Much work can be saved and devices of less precision used in the average computer if full advantage is taken of the simplification in instrumentation that may be effected by juggling the fundamental computer equations. While all the methods of mathematics are possible sources for these simplifying steps, a few of the more important methods can be listed specifically.

Among these is the principle that a problem may often best be solved in terms of increments from a chosen set of values rather than in terms of the full magnitude of the quantities involved. Thus, for example, in a computer converting Loran data to rectangular coordinate data covering a small region, it would be desirable to work in terms of distances from an origin at the center of the region. This has the advantage that devices with fixed percentage errors are operating on quantities whose magnitudes are relatively small.

A frequently useful method of simplification is that of zero shifting. Where both positive and negative values of data must be handled, a simple renumbering (shift of zero) will allow the same data to be represented on a data scale that does not include negative values. There are, obviously, many cases where this procedure will not work. Addition of constants (shift of zero) has been used in multiplying devices to avoid operating with zero inputs to the devices, even though the data go to or through zero. A discussion of this point is given by Fry.[1]

A common mathematical procedure of considerable use in computer design is that of series expansions and approximation. This method is particularly helpful when complicated nonlinear functions must be represented. The possibility of series expansions or approximations should at least be considered carefully when such functions are encountered.

For problems in which a point is represented by simple coordinates, rectangular, cylindrical, and spherical coordinates should be considered, and the most appropriate selected. In almost every case one of the three will be found to have a distinct advantage over the others. The two computers of Chap. 7 represent examples of position representation in rectangular and spherical coordinates, respectively. Other types of

[1] General references on computers include the following: F. J. Murray, *The Theory of Mathematical Machines* (good bibliography), Kings Crown Press, New York, 1947; M. Fry, "Designing Computers," reprinted from *Machine Design*, August 1945 to February 1946, Penton, Cleveland, 1946; H. Ziebolz, *Analysis and Design of Translator Chains*, Askania Regulator Co., Chicago, 1946; A. C. Blaschke, "Solution of Differential Equations by Mechanical and Electromechanical Means," AAF Eng. Div. Report TSEPE-673-4, 1946, available as Dept. Commerce No. PB 10298 (bibliography has 108 listings); and V. Bush, and S. H. Caldwell, "A New Type of Differential Analyser," *Jour. Franklin Inst.*, **240**, 255, October 1945.

coordinate systems may also be considered, of course, but rectangular, cylindrical, and spherical coordinates are by far the most common types used in computers.

In generating nonlinear functions, a frequently used technique involves the generation of an approximation to the desired function by some very simple and reliable device together with a nonlinear element (such as a cam) supplying only the correction to the approximation rather than the entire quantity being generated. Cams are also used in this manner to remove errors when the simple device referred to above generates the function desired, but with insufficient accuracy or with errors introduced by external conditions. A good example of this type of cam correction of errors is found in the Bendix-Pioneer Flux Gate Compass Master Indicator. Another simplification technique, that of implicit functions, is treated separately in Sec. 2·5. It is probably the most important of the methods available to the computer designer.

2·4. Explicit Analogue Computers.—As has been mentioned in the previous chapter, analogue computers operate by identifying variables in one physical system with those of a different physical system obeying equations of the same form but usually with different constants. It is this simple and fairly obvious principle which forms the basis for most of the computers that are discussed in the present part of this volume. When the fundamental equations cannot be written simply or instrumented readily, implicit analogue techniques may be required. These are discussed in the following section.

2·5. Implicit Function Techniques.—Implicit function techniques are among the most important tricks used in computer design. Through the use of the implicit function techniques described in this section and elsewhere, considerable simplifications and increased accuracy are easily achieved in many typical computers.

TABLE 2·1.—CORRESPONDING IMPLICIT AND EXPLICIT EQUATIONS

	Explicit	Implicit
Subtraction	$z = x - y$	$z + y = x$
Division	$z = \frac{x}{y}$	$yz = x$
Integration	$z = \int y\,dt$	$\frac{dz}{dt} = y$
Square root	$z = \sqrt{x}$	$z^2 = x$

There are numerous computations for which direct physical analogues are inconvenient, insufficiently accurate, or not available but for which another method of solution is possible. This other method is similar to the mathematical operation of computing the function $z = g(x,y)$ from the defining relation $f(x,y,z) = 0$; that is, of deriving an explicit function from the corresponding implicit function. It is especially convenient when the function $f(x,y,z)$ has a relatively simple instrumentation and the function $z(x,y)$ is difficult to instrument. It is used, for example, to perform subtraction by means of addition, division by means of multiplication, integration by means of differentiation, and the extraction of square roots by means of squaring. The explicit and implicit equations corresponding to these operations are given in Table 2·1.

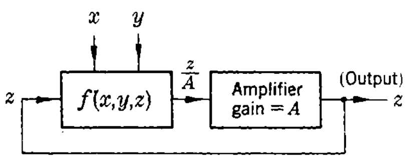


FIG. 2·1.—Solution of implicit function using feedback high-gain amplifier.

The simplest method in principle of producing the function $z(x,y)$ from the equation $f(x,y,z) = 0$ is to use a high-gain amplifier with feedback as shown in Fig. 2·1. The variables x, y, and z are fed into a device producing the function $f(x,y,z)$; the output is amplified by a factor A and fed back as the variable z. The result, if the system assumes a stable state, is that the device solves the equation

$$f(x,y,z) = \frac{z}{A}. \tag{1}$$

If A is sufficiently high,[1] the device produces values of z such that

$$f(x,y,z) \approx 0. \tag{2}$$

The quantity z/A may be called an "error signal." Thus there is an inherent error Δz in z which may be expressed by

$$\left(\frac{\partial f}{\partial z}\right)_{z=0} \Delta z = \frac{z}{A},$$

or

$$\Delta z = \frac{z}{A\left(\dfrac{\partial f}{\partial z}\right)_{z=0}}. \tag{3}$$

Consider, for example, the case of division. This may be accomplished either by a device that divides directly or by a device that multiplies, used with feedback. In performing any other operation a similar set of alternative methods, one with and one without feedback, is often possible. A choice between the two alternatives is made on the basis of the usual engineering considerations, such as accuracy, cost, availability, complexity, weight, size, etc. A schematic diagram and a possible instru-

[1] In Vol. 18, design procedures are given that are suitable for stabilizing high-gain amplifiers.

mentation for division using multiplication and feedback are shown in Fig. 2·2. The input y multiplies the input z. The product yz is then subtracted from the input x. Thus in this case $f(x,y,z)$ is $x - yz$, and

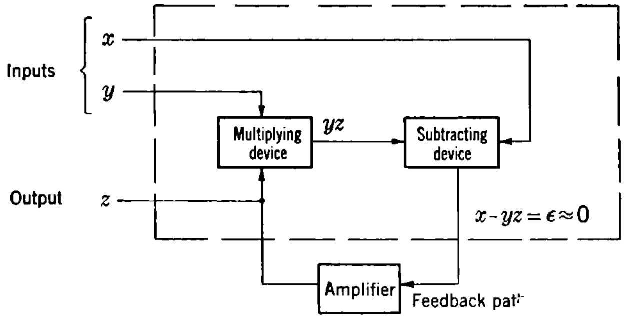


FIG. 2·2.—Block diagram of computer for $z = x/y$ using $yz = x$.

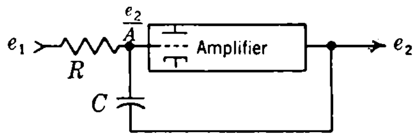


FIG. 2·3.—Integrating circuit.

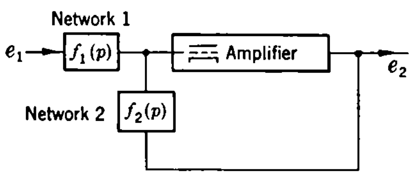


FIG. 2·4.—Feedback circuit for general operational equation.

the explicit function produced by equating this to zero is $z = x/y$. The inherent error is

$$\Delta z = \frac{z}{A\dfrac{\partial f}{\partial z}} = \frac{z}{A(-y)} = \frac{-x}{Ay^2}. \tag{4}$$

An application of this method to the solution of the equation

$$y^2 + \sin y + x = 0$$

is discussed in Sec. 6·1.

This method may also be used for producing the negative of a voltage (Chap. 3) or for integration with respect to time (Chap. 4). An integrator circuit is shown in Fig. 2·3. There is again an inherent error due to finite gain, best expressed (Chap. 4) in a different way from the above expression. The same method can be used with more complicated integro-differential functions. The general circuit of Fig. 2·4 produces an approximate solution of the equation

$$f_1(p)\mathfrak{L}(e_1) + f_2(p)\mathfrak{L}(e_2) = 0,$$

where $f_1(p)$ = transform expression for admittance of Network 1,
$f_2(p)$ = transform expression for admittance of Network 2,
$\mathfrak{L}$ = the Laplace transformation.[1]

In high-gain feedback circuits, special care must be given to the prevention of oscillations. The higher the gain the more difficult this becomes. The general procedure is to design networks that give the desired shape to the loop gain and phase-shift characteristics.[2] The design procedure in this respect is similar to that for servos as discussed in Chaps. 9 to 11.

Feedback with Integration.—Where feedback is used in implicit solutions, it is often of value to modify the error signal by other than straight amplification. Shaping loop gain and phase characteristics to achieve stability is one such modification that has already been mentioned. A form of such modification is integration of the error signal. If, for example, the error signal is used to operate a simple servomechanism whose motor forms part of the complete feedback loop and whose speed is proportional to the error signal, then integration of the error signal with respect to time is achieved. This has the useful effect of eliminating (in practice, of reducing) steady-state position errors. This subject is discussed in more detail in connection with servo theory in Chaps. 9 to 11 and by Hall.[3]

Additional integrations or differentiations may also be used to take fullest advantage of the data available rather than merely to achieve stability when high gain is used. The stability modifications must, of course, be made after the integration, etc., modifications have been selected.

Feedback without High Gain.—For some computations it is convenient to use a circuit that is somewhat like that of Fig. 2·1 except that a high-gain amplifier is not present. Such a circuit or device is shown schematically in Fig. 2·5. The equation that it solves is

$$f(x,y,z) = z. \tag{5}$$

[1] See Chaps. 9 to 11. In these chapters, the symbol s rather than p is used to represent the complex variable.

[2] H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945; and Vol. 18, Radiation Laboratory Series.

[3] A. C. Hall, *Analysis and Synthesis of Linear Servomechanisms*, Technology Press, Massachusetts Institute of Technology, 1943.

The conditions under which this method may be used are that a solution exists and that the desired output z may be expressed as a function of x, y, and z, this expression being of a form suitable for instrumentation with available techniques. If $f(x,y,z)$ contains an additive term z with unity coefficient and its computation involves one or more unilateral elements, then the feedback of z to the point in the $f(x,y,z)$ computations where it is added to the other terms must be unilateral; in other cases it may be bilateral. For good operation and high accuracy, $\partial f/\partial z$ should be much larger or much smaller than unity.

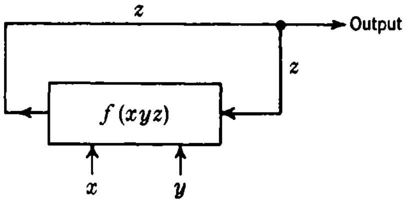


FIG. 2·5.—Feedback circuit without high-gain amplifier.

The equation

$$e^z + 3z + y + 2x = 0 \tag{6}$$

is readily changed to either of the following forms,

$$\left.\begin{aligned} e^z + 4z + y + 2x &= z \\ \frac{e^z + y + 2x}{-3} &= z \end{aligned}\right\} \tag{7}$$

and may be used as such provided the left-hand sides of the expressions may be instrumented. The equation

$$e^z + y + 2x = 0 \tag{8}$$

may be changed to the proper form by adding a z to each side, becoming

$$e^z + y + 2x + z = z \tag{9}$$

which requires unilateral z feedback if any of the operations on the left hand side are unilateral. Equation (7) can also be rearranged to such forms as

$$\ln(-2x - y - 3z) = z, \tag{10}$$

which may be used provided the left-hand side of the expression can be instrumented.

A comparison of the two types of feedback connection is of interest. When put into the proper form for use in a feedback loop without high gain, the resulting equation may be more difficult, as difficult, or easier to instrument than the same equation in optimum form for the more conventional feedback with amplification. The comparative difficulty of instrumentation is always an important factor in the choice of a method. When appreciable time constants are associated with the circuits or devices used, the feedback amplifier method assures a faster transient response than the no-amplifier method. Usually when the no-feedback amplifier method is used, some means is needed of compensating for decrease of scale factor due to the instrumentation of the various terms of the left-hand side of the equation of the form of Eq. (5). Where electronic circuits are concerned, this may require a stable low-gain amplifier, whereas in the feedback amplifier method a high-gain amplifier with little restriction on gain variations will suffice. For these reasons, a choice between the two alternative amplifiers may often favor the latter, although the former is simpler. In mechanical systems, mechanical amplifications are similarly to be considered, although here the problem of gain stability is not so serious as in the case of electronic amplifiers.

Interchange of Variables.—There are several ways in which the inverse of a function can be produced without the use of feedback. Probably the simplest is possible in the case of a mechanical function-producing device having two shafts that may be used alternately as input and output. One device of this sort is the combination of a cone and a cylinder[1] by a wire that winds onto one as it unwinds from the other; this device can be used to produce either squares or square roots. The interchange of mechanical inputs and outputs is limited by the nature of the function produced; if in some region the derivative of output with respect to input is nearly zero, the variables cannot be interchanged in that region.

Another type of interchange of variables occurs in the case of some two-terminal electrical devices with special voltage-current characteristics. Ordinarily the voltage is considered to be the independent variable and current the dependent variable, as in the case of diodes, triodes, and crystals. If a current generator rather than a voltage generator is used, however, the inverse function may be produced. This procedure is used in the logarithmic multiplying device of Chap. 5, in which the exponential grid-current characteristic of a triode is used with a high grid resistor to produce the inverse function, a logarithm.

A third method of interchanging variables finds particular application to computation with periodic waveforms. If a repeated voltage waveform is compared by means of an amplitude comparison device (Vol. 19, Chap. 9) with an input voltage, a pulse or other indication of the time of coincidence may be obtained. If the original waveform is a given function of time, this same function of a slowly varying input voltage may be obtained by time selection (Vol. 19, Chap. 10). Thus, for example, a parabolic waveform can be used to produce either squares or square roots.

The principle of repeated functions with amplitude comparison to compute inverse functions or with time selection to obtain direct func-

[1] Sec. 5·11.

tions may be applicable in a number of ways to a wide range of problems, although its use is not widespread. Its principal disadvantage, apart from considerations such as complexity, is that of the time required for the repetition period of the repeated function. An interesting application of this principle is found in the Keinath sweep balance multiple-data recorder.[1]

Stability in Implicit-function Loops.—Some general observations may be made concerning the realizability of solutions in loops such as that of Fig. 2·1. It is necessary that in the case of constant input the polarity of feedback be such as to produce equilibrium at the desired solution. This condition is equivalent to stating that the polarity of the amplifier must have the opposite sign from $\partial f/\partial z$ in the desired region. If the function $f(x,y,z)$, assumed to be continuous, has multiple real roots for $x = x_1$ and $y = y_1$, some of these roots may have $\partial f/\partial y$ of the proper polarity to make the roots stable and others may not. An example of this case is shown in Fig. 2·6. The arrows indicate the direction in which the system will go; the central root may be considered a position of unstable equilibrium. If the polarity of feedback is reversed, the formerly stable roots will become unstable, and vice versa. When the system is turned on, its behavior will depend on the polarity of feedback, on the region in which the system was when turned on, and on the changes that take place during warm-up. If the polarity of feedback is such as to make the outermost roots unstable, the system may oscillate at the limits of its region of operation. In order to examine stable inner roots it is necessary either to restrict the region of operation by modifying the device or to force it into the region desired, where it will then assume a stable state.

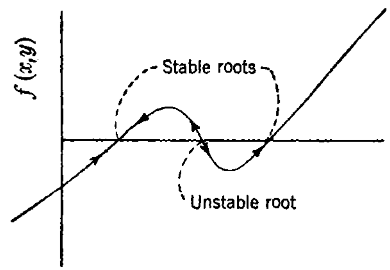


FIG. 2·6.—Stable and unstable roots.

A principle that is often useful in computer design may be generalized from the "rule of torques" set forth by Bush and Caldwell.[2] This generalized principle may be stated as follows: Every input to every element or "block" in a computer must be actuated by one and only one driving source.

2·6. Tolerances and Errors.—A standard designation for errors has been followed by the Navy for fire-control computers. According to

[1] G. Keinath, "The Keinath Recorder," *Instruments*, **19**, 200, April 1946.

[2] V. Bush and S. H. Caldwell, "A New Type of Differential Analyser," *Jour. Franklin Inst.*, **240**, 255, October 1945.

this designation, errors are classified as Class A, Class B, and Class C. Class A errors represent the deviation of the operation of the equipment from the theoretical design operation. They include reading errors, incorrect adjustments, errors in components, etc. Class B errors represent the mathematical approximation of the design. The sum of Class A and Class B errors is called a Class C error and represents the *accuracy* of the equipment. Class A errors when reduced to a minimum by accurate adjustment and accurate reading, etc., represent the *precision* of the equipment. Where Class B errors are required, the distinction between these terms is a useful one. With the exception of Chap. 5, most of the discussions of Part I do not involve Class B errors and therefore the terms "accuracy" and "precision" are used interchangeably.

In considering the accuracy of computing equipment, the question of calibration is encountered, particularly in the case of electronic circuits. In order to specify completely the accuracy of an equipment, its errors should be specified in terms of the frequency of calibration or of time since last calibrated. The principal reason for this is that components of all types change with time and use. The process of component drift may be very complicated and usually involves such things as thermal and moisture cycles, mechanical wear, creep, vacuum tube cathode disintegration, vibration, chemical reaction, etc., but regardless of the explanation, some drift is usually to be expected, and calibration intervals must be chosen so that errors due to component drifts are not excessive.

The terms component tolerance and component variation are frequently encountered. *Tolerance* is an allowable deviation as from a specified value under fixed conditions and is used to define permissible errors of manufacture or construction. Thus, a resistor might be partially specified as a 100-k 5 per cent resistor, the 5 per cent being the manufacturing tolerance. Component *variation* may be defined as the change of a component value as conditions change. Thus the 100-k 5 per cent resistor might have a value of 103 k at 20°C and change to 104 k at 70°C. The 1-k change would be referred to as the variation in this case.

In general, the effect of errors introduced by: component tolerances and variations from design center values, reading errors, calibration errors, etc., is treated by considering each component or other error source independently, computing the effect on the over-all operating accuracy to be expected from each of the given errors, and combining results by simple addition to give a limiting error or by probability methods (to be discussed) in order to arrive at a probable error. There are many situations, however, in computer design, where this type of analysis leads to false conclusions. As is described in the following section and elsewhere, variations in several components may often be made to cancel each other. For example, a divider network composed of elements with matched temperature coefficients operated at the same temperature will show no change in the division ratio with changes in temperature. Such components, whose errors are not independent, must be analyzed together.

The subject of the combination of errors of independent components, groups of components, or other independent error sources is one that can be approached from two different viewpoints, each of which may be applicable under certain circumstances. In many cases, the only satisfactory procedure is to design the computer such that even if all the effects of tolerances and variations under expected operating conditions add unfavorably, the devices will still operate within the required limits of accuracy, assuming that the desired performance can be described in terms of limiting error. In other cases, particularly where the use of the computer itself can be treated on a probability basis, there is considerable justification for combining errors according to the methods of probability. For example, such an approach would be justified in the case of a bombing computer, since bombing is usually evaluated on a probability basis. A serious problem immediately arises when this method is applied to electronic components. One must assign probable errors to components that are usually specified only in terms of limiting errors. A method that has considerable use as an approximation for engineering calculations assigns a "probable error" to a component equal in value to roughly one-third the limiting error. The term "probable error" is, of course, not strictly valid here; but since all the calculations are carried on as if it were a true probable error, the term will be used, with the understanding that it involves a big approximation. In a normal distribution, the figure of three times the probable error would include 96 per cent of a large number of samples, whereas in this approximation it includes all samples. It must be emphasized that the results of an analysis of this type cannot be more accurate than the assumptions on which it is based. However, the method allows calculations to be made that are of great assistance in engineering a design where limit values added unfavorably would give a completely misleading picture of the computer operation.

2·7. The Use of Servomechanisms in Computers.—A very important tool in computer design is the proper use of servomechanisms. A servo or servomechanism has been defined by Hazen[1] as "a power amplifying device in which the amplifying element driving the output is actuated by the difference between the input to the servo and its output." A detailed discussion of servomechanisms will be found in Part II of this volume and in Vol. 25. For purposes of illustration, a simple example of the use of a servo in a computer is shown in Fig. 2·7. This simple

[1] See Chap. 8.

computer is the same bridge circuit of Fig. 1·1, but with the null meter replaced by a servoamplifier and servo motor which adjusts resistance Z such that the bridge is always balanced, i.e., such that the voltage between c and d is zero. Its use in this simple computer allows the computer to be automatically operated rather than requiring an operator to adjust the resistance Z.

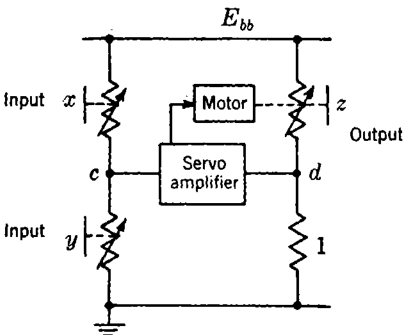


FIG. 2·7.—Bridge computer for $z = x/y$, illustrating use of servomechanism for automatic balancing.

The use of servos in computers allows a designer to apply high-gain feedback techniques to loops involving electromechanical elements, whereas without servos feedback technique is limited principally to all electronic loops. Servos are also useful as parts of data-transmission systems. A major advantage of servomechanisms is included in the definition, namely, that of torque amplification. The property of torque amplification allows blocks to be connected in such a way that they do not act back to disturb previous blocks.

2·8. Error Cancellation.—Accuracy is, of course, a factor of great importance to the computer designer. Accuracy may be achieved by "brute force" methods, in this case by making all components and operations accurate to the requisite degree. Accuracy may also be achieved by using components whose errors may be relatively large but that are used in such a way that their errors cancel or nearly cancel each other. Error cancellation is a concept of great fundamental importance to the computer designer, and many applications of this concept are found in every well-designed computer. Through the use of applications of this concept, computers are made simpler for the same accuracy or more accurate for the same complexity as compared with designs wherein accuracy has been achieved merely by reducing all individual errors independently. Although not always evident, the basic principle used in most error cancellation methods is that of the bridge circuit.

Following are listed a few common applications of the techniques of error cancellations. The list of applications discussed, although it is not comprehensive, should allow other applications of the principle to be easily recognized.

Reference and Power-supply Voltages.—When precision is an important factor, it is necessary to ensure that variation of power-supply and reference voltages will introduce only small errors. As described above, two approaches to this problem are made. In the first and most obvious, but usually the most costly method, the voltage and power supplies are highly regulated so that accurate and stable voltages are available. In practice, the limiting factor in this procedure is usually the standard to which voltages are regulated. This subject is discussed in detail in Chaps. 15 and 16.

The second approach is that of cancellation of errors. Typical methods falling in this category are bridge circuits such as the one shown in Fig. 2·2, change of sweep slope in time-modulation circuits to cancel changes in "pick-off" or comparison voltages (Sec. 3·17), and condenser tachometer voltage output variation with supply voltage to cancel changes in velocity servo control voltages (Sec. 3·17).

Even when error cancellation methods are used to avoid the effects of supply variations, it is usually necessary to incorporate fair regulation in the voltage supplies. The reason for this is that although first-order cancellations can be taken care of by other means, second-order effects may remain, and these, together with transient errors, may introduce appreciable inaccuracies if voltages are allowed to change widely.

Where both a-c and d-c voltages are used in a computer as accurate representations of the magnitudes of quantities entering into the computation, it is desirable, following the principle of error cancellation, to regulate the alternating from the direct current or the direct from the alternating current. The principal argument for regulating the direct current from the alternating current is that to do the opposite by electronic means requires control elements such as vacuum tubes or saturable reactors that may introduce harmonics into the alternating current or require heavy and complex filters if appreciable power is taken from the a-c supply.

Other examples of error cancellation are the cancellation of altitude error in the so-called H plus B bombing computers (*cf.* Sec. 3·7), cancellation of resistance variations in a voltage divider by matching temperature or other coefficients (*cf.* Sec. 3·2), matching error curves of synchros to reduce operating inaccuracies when two synchros are operated together, etc.

A careful consideration of any computer design from the standpoint of possible applications of error cancellation methods is generally profitable.

2·9. Data Smoothing, Speed of Operation, and Stability.—It is the rule rather than the exception that the input data of an analogue computer either have power-frequency spectra differing from that which would be considered optimum or have extraneous signals (usually called "noise") mixed with them. Often input data are subject to both defects. The average computer, therefore, will incorporate means for modifying the power-frequency spectra of the data as they pass through the computer. Usually this modification takes the form of "smoothing," that is, removing or reducing high-frequency components. This is illustrated by the action of an antiaircraft computer working with automatic-tracking-radar line-of-position data. Such radar data, as furnished by an equipment such as the SCR-584,[1] contain a considerable amount of high-frequency "noise." A computer, such as the BTL Mark IX director, used with this data must do considerable "smoothing" in the process of computing data for gun pointing in order to assure a high probability of hits. Different quantities must be smoothed in different ways; for example, rate data derived from position data must be "smoothed" considerably more than the position data. Smoothing is not the only type of frequency modification used. Removal of drifts, that is, reduction or removal of low-frequency components, is also occasionally necessary. The subject of modification of frequency characteristics is discussed in Chap. 11, with some introductory material in Chaps. 9 and 10.

The somewhat loosely used term "speed of response" is closely related to steady-state frequency characteristics. The relationship is analogous to the relationship between a function of time and its Laplace transform. Speed of response is usually treated more satisfactorily from a quantitative standpoint in terms of frequency characteristics. Velocity and acceleration errors are also useful quantitative indices of "speed of response" and are also discussed in Chaps. 9 to 11. Particular reference is made to the Farrell method for determining velocity and acceleration errors from decibel loop gain vs. log frequency plots (Sec. 10·4), and a simple criterion for the choice of transmission vs. frequency characteristics reported by Graham (Sec. 11·10). This criterion specifies that the transmission characteristic of a servo or computer should approximate the amplitude spectrum of the useful part of the most probable input signals, i.e., exclusive of noise, which is assumed to be uniform with frequency.

Where feedback loops are involved, smoothing circuits and devices must be carefully considered from the standpoint of stability. Reference is made to Chaps. 9 to 11 for the theory of stability in feedback loops.

2·10. Reduction of Weight, Size, and Power Dissipation.—There are a number of points in good engineering design that, if followed carefully, may substantially reduce the weight, size, and power dissipation of an electronic or electromechanical computer. While the subject of design and construction practices is discussed in detail in Chap. 19, several of

[1] "SCR-584 Radar," *Electronics*, **18**, 104–109, November 1945; 104–109, December 1945; and **19**, 110–117, February 1946.

the more important points will be mentioned here, since they are important basic principles in computer design. Size, weight, and power dissipation are intimately related, and circuit changes resulting in a reduction in any of the three items may often be reflected as reductions in the other two.

Some important design principles are

1. Scale factors in a circuit or device should be such that some operations are carried out at small signal amplitudes and followed by a signal amplification in a stage just preceding elements such as detectors that require large signal amplitudes for good percentage accuracy. This saves power.
2. Wherever possible, tubes should be normally nonconducting. This also saves power.
3. The use of miniature and subminiature tubes and some other special components saves both space and appreciable heater power as compared with standard-size tubes.
4. The use of special card or subassembly methods (*cf.* Chap. 19), short leads, and sometimes sealed enclosures allows circuits to be operated at high inpedance, thus saving plate and other power and reducing inaccuracies due to loading of previous circuits.
5. Where feasible from the standpoint of accuracy, etc., it is simpler and more compact to compensate components for temperature changes rather than to regulate temperature.
6. By close design it is possible to cut down the large factors of safety usually provided as a substitute for accurate knowledge of components and working conditions.

REPRESENTATION OF QUANTITY

2·11. Fundamental Concepts.—The term representation of quantity as used here refers to the physical property of an analogue computing system that is identified with a specific quantity in a system for which computation is required. Thus, for example, in an aircraft navigation computer, a specific quantity, altitude, in the system for which computation is desired may be identified with an a-c voltage in the analogue system that is to perform the computation. In considering the representation of quantity it is useful to distinguish between single-scale and multiscale devices. Although multiscale techniques can be used in addition and subtraction, they are not usually suitable for other computing operations.

A *single-scale* data system may be defined for most purposes as a system in which the full working range represents the full range of the variable represented. An ordinary potentiometer is a simple example of a single-scale device. A *multiscale* data system is one in which two or more channels are provided, one channel in which the full working range represents the full range of the variable represented and one or more channels in which the full working range represents only a portion of the full range of the variable represented and hence is repetitive in covering the full range of the variable represented. The first channel is called, alternately, the *coarse*, *single-speed*, or *one-speed* scale, and the latter channels are called the *fine* or *high-speed* scales and provide a "magnification" of the coarse scale. A high-speed channel may be used without a coarse or single-speed channel, but in this case an ambiguity of indicated value will exist. The principal advantages of multispeed data systems are increased accuracy, increased resolution, and increased gain. These are obtained at the expense of increased complexity.

Three criteria are particularly important in evaluating a specific type of representation for a specific quantity at a specific point in a computer. These considerations are impedance, scale factor, and useful range. The question of whether bidirectional data or unidirectional data are used is included in the third item. Impedance is considered from the standpoint of the effect that a device will have on a prior device and its susceptibility to pickup. The question of scale factor is a most important one, particularly in d-c computers, since scale factors must be large enough to make the proportional change of tube characteristics small yet must not be so large as to necessitate unduly large tubes, to cause nonlinearity trouble with normal tubes, or to waste power and add unnecessary weight. The useful range of a device must be considered simultaneously with the scale factor in evaluating a representation of data. Useful range is limited by such factors as nonlinearity, power dissipation, mechanical limits, voltage limits, etc. Some devices need operate only on the magnitude of a quantity, whereas in other devices both magnitude and sign must be taken into account. For example, altitude is a quantity that is represented by its absolute value, whereas latitude may be either north or south and hence must have a sign associated with its magnitude.

2·12. Seventeen Important Types of Data Representation.—The following list of representations covers those which are most frequently found in electronic and electromechanical computers, although the list is by no means complete. Following each type of representation will be given the abbreviation that will be used in subsequent chapters in classifying the inputs and outputs of the standard blocks from which a computer may be built. In this system of classifying abbreviations, the symbols preceding the colon represent the inputs; those following the colon, the outputs. Thus, a resistor dividing circuit having voltage and resistance inputs and a voltage output would be represented by the notation $E, Z{:}\ E$. It is to be noted that classification by data representation is lower in rank than classification by operation performed, according to the system used in Part I of this volume. The seventeen important types of data representations are

Force, F.
Pressure, p.
Torque, T.
Translational displacement, S.
Angular rotation, θ.
Translational velocity, V.
Angular velocity, ω.
Translational acceleration, a.
Angular acceleration, α.
Voltage, E.
Current, I.
Charge, Q.
Impedance, Z, usually resistance.
Frequency, f.
Phase, ϕ.
Count, N.
Time interval, t.

2·13. Some Characteristics of Various Types of Representation.—It is not practical to compare every type of representation with every other type. A few generalities can, however, be pointed out. It must be remembered that there are exceptions to many of the following statements, and each device should be considered individually.

Electrical vs. Mechanical.—The advantages of mechanical devices are: fewer adjustments; no vacuum-tube or other similar component drifts; greater ease in understanding; and less difficulty in writing complete specifications.

The disadvantages of mechanical devices are: they are subject to friction and wear; complicated procedures of design and construction are involved; and highly skilled labor is needed for producing precision devices.

Advantages of electronic devices include: flexibility, in that many devices may be made from standard parts; cheapness, in that semi-skilled labor can wire complicated equipment and expensive tools and dies are not required; short design time; in some cases, accuracy; lightness; and speed of response.

Electronic disadvantages are: more controls are required in precision circuits; vacuum-tube drifts and burnouts are sometimes bothersome (this is normally not a serious limitation if proper design precautions are taken): and electronic devices are usually more complicated to understand and service, although new techniques embodying subassembly construction are tending to change this.

Alternating vs. Direct Current for Data Representation.—Alternating-current advantages are the following: vacuum-tube zero drifts do not seriously affect the over-all accuracy; transformers can be used; in rotating machinery no commutators are required; d-c voltage levels may be more easily isolated when a-c data are used; and additional variables are available in the form of frequency and phase.

Alternating-current disadvantages are: difficulties arise in differentiating and integrating; a-c tachometers are not very satisfactory; and problems of phase shift, frequency variation, and unwanted electrostatic and magnetic pickup are generally troublesome.

Direct-current advantages are: differentiating and integrating are easy; because of this, simple and reliable phase lead circuits can be used; no phase shift or frequency problems are encountered; and no electrostatic or magnetic pickup is involved.

Direct-current disadvantages are: drifts in vacuum tubes cannot be distinguished from d-c data; contact potential troubles are encountered; difficulty with d-c voltage level is frequently a severe design limitation; and no good d-c resolver is readily available.

Time Interval and Pulse Waveforms.—These are of some interest but are usually more complex than other alternatives. They are most likely to be useful when the output data representation required is a time interval.

Impedance.—The impedances generally used in computers are resistors or transformers and resolvers. Resistances in the form of variable resistors or potentiometers are good, simple transitions from mechanical to electrical representation of data and are particularly useful in bridge circuits.

Phase Shift.—In some cases phase shift proves to be a reliable and accurate means of representation. Phase-shift techniques, however, require special design care; and where high accuracy is required, close frequency tolerances must be imposed. Phase-shift techniques are widely used in accurate time-modulation and -demodulation equipment (*cf.* Vol. 20) which may occasionally be combined with computing equipment.

Count.—Count is of more importance in the digital computer field than in the analogue computer field. In the latter field, a typical promising application is in process control computers where the input data are the number of items passing a counter station.

Frequency.—Frequency is more likely to be used as a data representation for data transmission (telemetering) than for use in a computer. A few uses in computers are common, as, for example, the use of a-c frequency to represent the speed of rotation of a shaft on which is mounted an alternator supplying the alternating current. It is worth remembering that phase is the integral with respect to time of frequency or, conversely, frequency is the time derivative of phase.

2·14. Summary of Chapter.—The following steps have been listed as a suggested systematic design procedure: preliminary information, block diagram, preliminary design, detailed performance analysis, detailed design, construction of model, and repetition of steps. Basic principles and techniques used in computers have been discussed under the following topics: simplifications in the formulation of computer equations and choice of over-all block diagram; explicit and implicit analogue computers; errors, tolerances, and variations; error cancellation; use of servomechanisms in computers; data smoothing and stability; and reduction of weight, space, and power dissipation. The important subject of data representation has been treated in some detail. Seventeen important types of data representation are listed, and some general comparisons of some of their characteristics have been given. A convenient system of abbreviated classification of operations from the standpoint of types of input and output data representations has been introduced and is illustrated in the chapter which follows.

CHAPTER 3

ARITHMETIC OPERATIONS

BY J. LENTZ AND I. A. GREENWOOD, JR.

3·1. Introduction.—Arithmetic operations described in this chapter include the operations of addition, subtraction, discrimination, multiplication, division, and the "identity operation" the latter term referring to changes in level, impedance, scale factor, or representation.

In presenting the highly varied material of this chapter, the authors have attempted to classify methods and devices according to the operations for which they are used rather than the method by which the operation is performed. Thus, for example, the method of multiplication that is based on the equation $xy = \int x\,dy + \int y\,dx$ involves both integration and addition but is classified as a method of multiplication, the end desired, rather than as a method of integration or addition, the means employed. The operations listed above are further subdivided according to the method used, and the shorthand notation discussed in Chap. 2 is used to indicate inputs and outputs.[1]

The operations described in this chapter may be said to be the heart of most electronic, electromechanical, and mechanical computing devices, for it is usually found that a generous proportion of the basic operations that must be carried out are those listed above.

3·2. Addition Using Parallel Impedance Networks (E **or** I:E **or** I).—It is well known that if a network of linear impedances is energized by two or more generators, the current or voltage at any specified point in the network can be expressed as the sum of the voltages or currents that each generator would produce were it alone connected to the network with each of the other generators replaced by its internal impedance. This property makes possible the use of such networks to perform the addition of voltages and currents. Subtraction of quantities as well as

[1] In this scheme, the symbols preceding the colon represent the inputs, those following it the outputs. Thus (E,E:I) would indicate a device with voltages as the two input representations and a current as the output representation. Usually only the most important variations of the representations used with any method are given, but it will be understood that others may be possible; for example, by including an appropriate resistor, most data represented as current can also be represented as voltage. Derivatives of the inputs and outputs given may usually themselves be considered inputs and outputs. Thus, a mechanical differential adding rotational displacements may also be used for adding rotational velocities, the output being also a rotational velocity.

the addition of quantities is also accomplished by this method by the representation of positive quantities by negative voltages or reversed currents. The following discussion is limited to a few of the simpler examples, as the general topic of addition of electrical quantities has been covered in some detail in Vol. 19, Chap. 18.

A simple illustration of this principle is shown in Fig. 3·1*a*, the operation of which is easily understood by reference to its Thévenin equivalent shown in Fig. 3·1*b*. This circuit has the advantage that each voltage source may have one terminal grounded, as is usual and convenient in vacuum-tube circuits. The network is particularly useful as an input network in a feedback amplifier where the algebraic sum of several voltages and a feedback voltage is used as the input error signal to the amplifier and is made nearly equal to zero by the feedback connection.

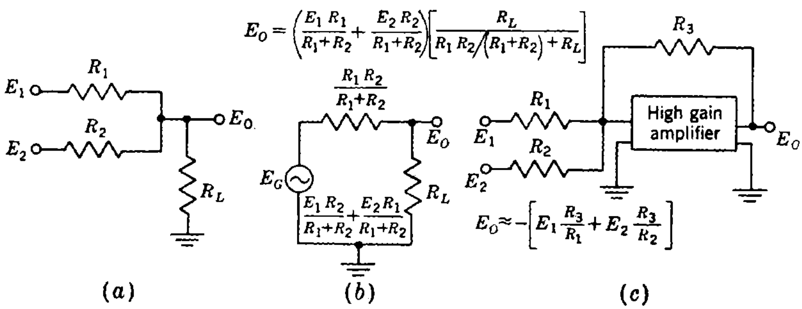


FIG. 3·1.—Addition with parallel impedance networks.

This is illustrated in Fig. 3·1*c*. With the circuit of Fig. 3·1*c* and assuming infinite gain in the amplifier, the output voltage E_o is proportional to the negative of the algebraic sum of E_1 and E_2, each multiplied by a constant that is a function of the network. There are many variations of this simple circuit, among the most important of which are the substitution of capacitances and inductances for the resistances R_3 and R_1 of Fig. 3·1, giving integrating and differentiating circuits. This subject is discussed in the following chapter and in Vol. 19, Chap. 18.

There are a number of practical considerations that should be taken into account in connection with this circuit. In order to obtain precision operation over a range of conditions, the ratios of the impedances must remain fixed. This means, among other things, that temperature coefficients must be matched to the desired precision, although there is usually no requirement that the absolute value of the coefficient be zero if all components operate at the same temperature. If, however, there are variations in temperature between components, such as might arise from local hot spots near high dissipation elements, then such temperature differences will act on the absolute temperature coefficients to produce errors. As an example, suppose resistors R_1, R_2, and R_3 of Fig. 3·1 are precision wire-wound resistors whose temperature coefficients are plus 300 ± 5 ppm/°C. In order to calculate maximum errors, assume that the coefficients of R_1 and R_2 are 295 ppm/°C and that the coefficient of R_3 is 305 ppm/°C. Then, for a 100°C change in temperature, E_o will become

$$E_o\big|_{100} = E_1 \frac{R_3(1.0305)}{R_1(1.0295)} + E_2 \frac{R_3(1.0305)}{R_2(1.0295)} = 1.00097E_o, \tag{1}$$

a change of only 0.1 per cent for 100°C change of temperature. On the other hand, if R_1 and R_2 are at the same temperature and R_3 differs from that temperature by only 10°C, then

$$E_o\big|_{10^\circ \text{ diff}} = E_1 \frac{R_3(1.00305)}{R_1} + \frac{E_2 R_3(1.00305)}{R_2} = 1.00305E_o, \tag{2}$$

a change of 0.3 per cent for 10°C temperature difference, or only 3.3°C difference for the same 0.1 per cent change calculated above for a 100°C change of temperature.

From this discussion it follows that elements used in circuits of this type should have low absolute temperature coefficients and should be mounted to ensure a minimum of local temperature variation. These requirements have sometimes necessitated use of a copper temperature-equalizing strip between elements where low absolute coefficients are unavailable or undesirable for other reasons but where high precision must be maintained.

If, as is usual, the operations performed by the circuit of Fig. 3·1 or those like it must be accurate, then either all voltages and all impedance ratios must be accurate, or either impedances or voltage scale factors or both must be made adjustable to allow for production tolerances in impedances or voltages. If the network has been made from large fixed resistors in order to avoid loading the voltage sources E_1 and E_2, then large adjustable resistors will be required. The stability and temperature coefficients of such adjustments must be carefully taken into account in the choice of network impedances and tolerances. The commercial availability of resistor networks matched in value to 0.1 per cent makes it possible in many applications to eliminate impedance ratio controls where resistance adding circuits are used. Alternately, other design considerations may dictate adjustment of voltage scales, using this adjustment to compensate for production tolerances in impedance ratios.

Zero adjustments in this circuit are best made in the amplifier itself, and for a discussion of this problem the reader is referred to Vol. 18.

Bridge circuits are also useful for addition and subtraction of voltages and currents. In a balanced bridge network Fig. 3·2 the mutual coupling between inputs is reduced to zero. The most useful form of this circuit is the equal-arm bridge network, as illustrated in Fig. 3·3. For this circuit,

$$E_0 = \frac{Z}{2}\left(\frac{E_1}{Z + Z_1} + \frac{E_2}{Z + Z_2}\right), \tag{3}$$

and the mutual coupling between E_1 and E_2 is zero. The input impedance at either input is Z, Z_1 and Z_2 being source impedances.

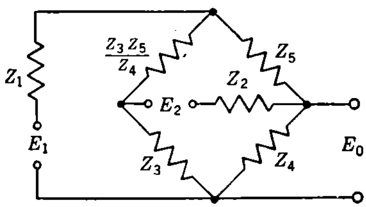


FIG. 3·2.—Balanced bridge.

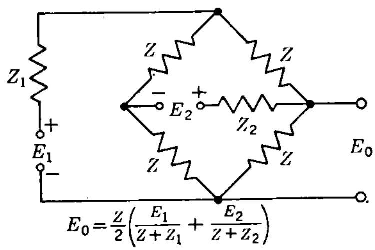


FIG. 3·3.—Equal-arm bridge.

3·3. Addition and Subtraction with Series Sources (E or I: E or I).—The fundamental principle underlying this method of addition and subtraction is identical with that given in the preceding section; namely, if a network of linear impedances is energized by two or more generators, the voltage or current at any specified point in the network can be expressed as the sum of the voltages or currents that each generator would produce were it alone connected to the network with each of the other generators replaced by its internal impedance.

If two generators of internal impedances Z_1 and Z_2 and internal voltages E_1 and E_2 are connected in series with a load impedance Z_L, the voltage E_L across Z_L is given by

$$E_L = \frac{Z_L(E_1 + E_2)}{Z_1 + Z_2 + Z_L}. \tag{4}$$

If E_1 and E_2 each represents a distance to a scale of, say, 1 ft per volt, then E_L represents the sum of the two distances to a scale of $(Z_1 + Z_2 + Z_L)/Z_L$ ft per volt.

It is to be noted that with series addition and subtraction, all generators but one must be such that their terminals may be isolated from ground ("float") if the desired sum or difference must be with respect to ground. If the sum or difference is desired as the voltage across two terminals that may both be isolated with respect to ground, then all but two generators must be isolated from ground. "Ground" is here used in the sense of a reference node; it is by no means uncommon for this node itself to vary with respect to ground as part of another circuit.

In using series addition and subtraction, the effects of stray impedances, usually capacitances, from various parts of the series network to ground must be carefully considered. Where alternating current is used, these stray impedances are apt to give troublesome phase and amplitude changes. If there is a-c information or ripple on the reference node

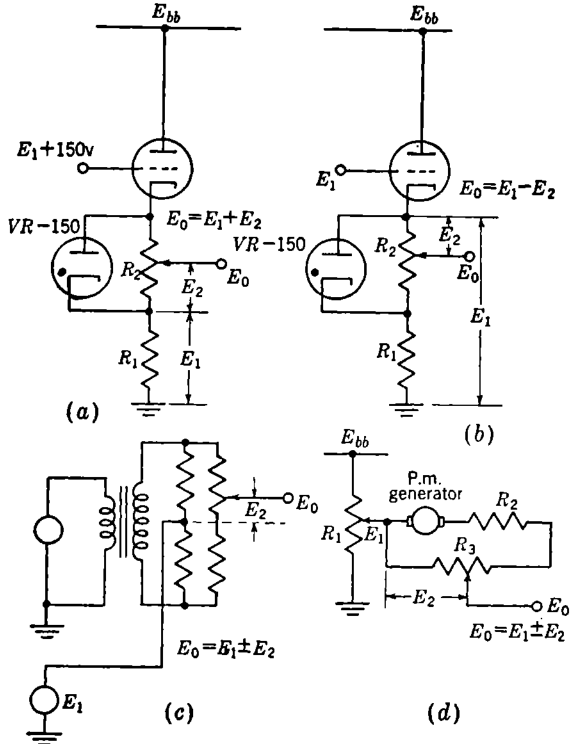


FIG. 3·4.—Addition or subtraction with series sources.

unbalanced impedances to ground may cause undesirable amounts of a-c voltage to appear in the series network.

A number of typical examples of series addition and subtraction are shown in Fig. 3·4.

3·4. Addition and Subtraction with Synchros ($\theta,\theta:\theta$ or E).—By the use of a chain of three synchros (Fig. 3·5), consisting of a synchro generator, synchro differential, and either synchro control transformer or motor depending on whether or not a servo drive is desired, a shaft rotation may be produced that is the sum or difference of two shaft rotations, and any of the three synchros may be remotely located with respect to the other two. Thus, in Fig. 3·5, $\theta_3 = \theta_1 + \theta_2$; or if connections or definitions of positive directions are changed, $\theta_3 = \theta_1 - \theta_2$ or $\theta_3 = \theta_2 - \theta_1$. The method may obviously be extended to a chain of more than three synchros.

This method has the advantage that multiscale methods may be used. By the use of, say, 36-speed synchros as well as the 1-speed synchros shown,[1] the accuracy of the 36-speed synchro chain may be realized in the addition or subtraction indicated. The method is therefore one of great accuracy and flexibility and is often used where the input and output data may be mechanical shaft rotations or may easily be changed to shaft rotations. A typical example of the use of this method is the

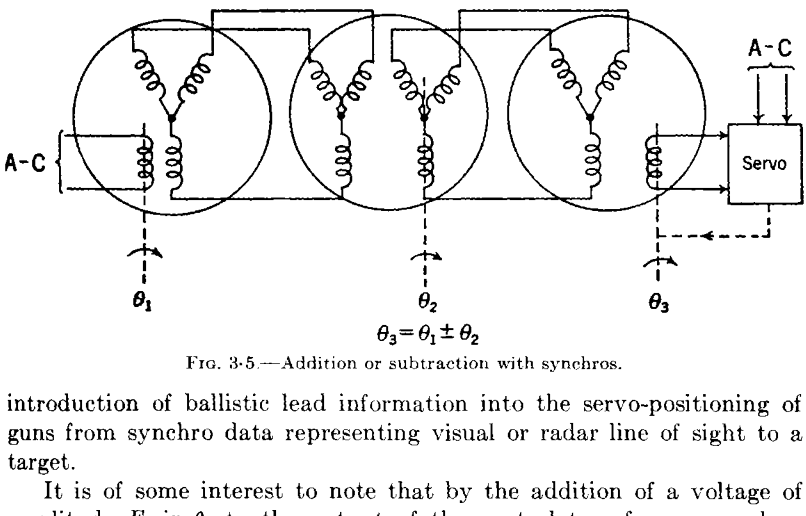


FIG. 3·5.—Addition or subtraction with synchros.

introduction of ballistic lead information into the servo-positioning of guns from synchro data representing visual or radar line of sight to a target.

It is of some interest to note that by the addition of a voltage of amplitude $E \sin \theta_4$ to the output of the control transformer synchro, where E is the maximum rotor output voltage, it is possible to shift the output shaft position from θ_3 to $\theta_3 \pm \theta_4$.

This method is useful when a shift of the output shaft by only a few degrees is desired and where electrical data are available. An example of the application of this method is the roll and pitch correction of a radar azimuth marking circuit using synchros by means of signals derived from a vertical gyro or gyros.

3·5. Impedance Addition and Subtraction $(Z,Z:Z)$, $(\theta,\theta:\theta)$.—It is frequently convenient to represent the physical parameters involved in the solution of a problem as impedances. Quantities represented by impedances may be added simply by connecting the impedances in series.

[1] See Sec. 13·8 for a discussion of multispeed data-transmission circuits.

Bridge methods of compounding mechanical rotations based on impedance addition have been found quite useful. In Fig. 3·6*a* is shown a type of system that has been employed to add two shaft rotations. Z_1 and Z_2 are two variable resistors whose resistances are linear functions of the rotations of their shafts. The bridge whose elements are Z_a, Z_b, $Z_1 + Z_2$, and Z_3, is kept in balance by means of a servomechanism, using as an error signal the voltage at the horizontal junctions of the bridge. The servo operates to adjust the impedance of the resistor Z_3. If Z_1, Z_2, and Z_3 are identical linear variable resistors, it is easy to see that

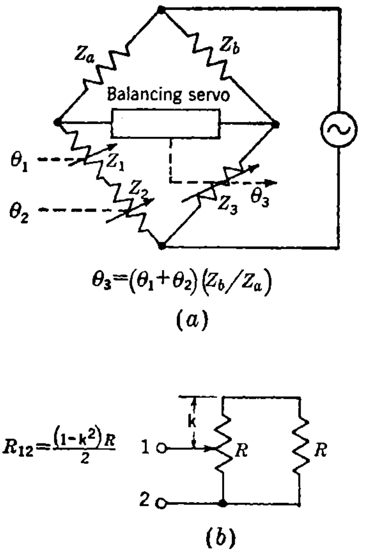


FIG. 3·6.—(*a*) Addition of shaft rotations with bridge; (*b*) element of squaring and square-rooting bridge.

$$\theta_3 = (\theta_1 + \theta_2)\frac{Z_b}{Z_a}. \tag{5}$$

By the use of a simple variation of this circuit, an interesting computing device may be designed to solve for one side of a right triangle, given the other two sides. The description of this circuit is presented here as a variation of the circuit previously described, although it performs a nonlinear operation on the sum of the squares of two quantities. It will be referred to again in Chap. 5.

Suppose that Z_1, Z_2, and Z_3 of Fig. 3·6*a* are identical, each being a potentiometer of resistance R bridged by an equal fixed resistance, as shown in Fig. 3·6*b*, and that $Z_b = Z_a/2$. If k is the fraction of the resistance of the potentiometer included between the upper terminal and the slider, then the resistance between terminals 1 and 2 is $(1 - k^2)\ (R/2)$; or if θ is the angle of rotation of the shaft and θ_t the total travel of the shaft, $k = \theta/\theta_t$ and $R_{12} = (1 - \theta^2/\theta_t^2)\ (R/2)$. When the bridge is in balance

$$\frac{1}{2}\left(1 - \frac{\theta_1^2}{\theta_t^2}\frac{R}{2} + 1 - \frac{\theta_2^2}{\theta_t^2}\frac{R}{2}\right) = 1 - \frac{\theta_3^2}{\theta_t^2}\frac{R}{2}, \tag{6}$$

or,

$$\tfrac{1}{2}(\theta_1^2 + \theta_2^2) = \theta_3^2,$$

$$\theta_3 = \frac{1}{\sqrt{2}}\sqrt{\theta_1^2 + \theta_2^2}. \tag{7}$$

If θ_1 and θ_2 are proportional to the perpendicular sides of a right triangle, then θ_3 is proportional to the hypotenuse.

In the practical design and construction of these circuits, variable resistors or potentiometers must be used whose total angle and total resistance or in some cases resistance change per degree may be specified to the desired accuracy. Alternately, suitable adjustments for variations in these units must be included. Simple analysis will show that the necessary adjustments are not easy to incorporate in all cases. It is also necessary to take into account the variable gain of the servo loop. The variation of servo loop gain is discussed in Sec. 11·9.

3·6. Addition and Subtraction with Mechanical Devices[1] (S or θ:S or θ).—The most frequently used mechanical device for addition and subtraction is the differential gear unit. This is so well known that no description is called for. The biggest problems with differentials have been backlash and procurement.[2] Good differentials require very careful workmanship and are likely to be rather expensive.

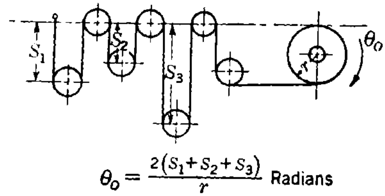


FIG. 3·7.—Addition using pulleys and tapes.

Mechanical motions are frequently added by means of hydraulic systems, in which the displacement of a piston is proportional to the sum of the displacements of two or more other pistons.

Another way of adding and subtracting mechanical motions is by means of levers. Nearly all the operations of computing may be performed with levers. Lever computers are discussed at length in Vol. 27.

Another mechanical method for addition and subtraction involves pulleys and tapes or wires and is shown schematically in Fig. 3·7. It produces an angular rotation θ_0 proportional to the sum of a number of linear displacements S_1, S_2, S_3 . . .

3·7. Addition of Time Delays (t,t:t).—This method appears suitable for addition but can probably be used for subtraction only through the use of implicit methods.[3]

If a series of time-modulation circuits are each triggered at the end of the previous delay, the total delay produced is, of course, the sum of the individual delays. This method has found its widest application in

[1] For an excellent summary of mechanical methods in computers, see M. Fry, "Designing Computing Mechanisms," reprinted from *Machine Design*, August 1945 through February 1946.

[2] Instrument differentials were made during the war by Ford Instrument Co., Librascope, International Business Machines, and others but may not now be available as units.

[3] See Sec. 2·5 for a discussion of implicit methods.

multiple-scale time-modulation circuits[1] and has also been used in the so-called "$H + B$" solutions of the triangle involving altitude, ground range, and slant range. In aircraft applications it is frequently more accurate to generate a slant range time marker corresponding to a certain (small) ground range by adding a quantity B to the altitude delay rather than by directly generating a time delay equal to the entire slant range. In the equation

$$B = \sqrt{R^2 + h^2} - h \tag{8}$$

B does not vary rapidly with h.

$$h + B = \rho, \tag{9}$$

where h = altitude,
R = ground range,
ρ = slant range,
B = difference between slant range and altitude.

The principal advantage of this solution lies in the fact that the same radar zero and setting errors that would appear in the generation and use of a slant range marker directly are canceled to a first approximation by their introduction into the h settings, providing the h marker is set to the first ground return (altitude signal) in the same manner that the ρ marker is set to a radar target. This allows the use of a short B delay whose absolute accuracy may therefore be good, added to an altitude delay whose calibration need only be good enough for computation of B. See Vol 22 for a more complete discussion of this method of solution of the altitude triangle.

The usual precautions regarding accurate, reliable triggering, independence of repetition rates, etc., should be observed.

3·8. Addition and Subtraction of Pulse Counts ($N,N{:}N$).—The method of addition and subtraction of pulse counts is the basis on which pulse digital computers, such as the ENIAC and EDVAC[2] computers of the University of Pennsylvania, are built.

Another application of pulse counting is in the Loran timer and receiver-indicator equipments.[3] In this equipment a pulse is fed back (injected) in a counting or dividing circuit, the amplitude of the injected

[1] See Vol. 20, Chap. 6.

[2] A number of articles on this computer development have appeared; see, for example, "Electronic Calculating Machine Is a Giant of Precision," *Elec. Mfg.*, **142,** April 1946; "ENIAC: War Dept. Unveils 18,000 Tube Robot Calculator," *Electronics*, **19,** 308, April 1946; and D. R. Hartree, "The ENIAC, an Electronic Calculating Machine," *Nature* (London), **157,** 527, Apr. 20, 1946.

[3] Volume 4 or any of a series of recent articles, such as "The Loran System," *Electronics*, **94,** November 1945; and "Loran Receiver Indicator," *Electronics*, **110,** December 1945.

pulse controlling the number of counts by which the firing period of the divider is reduced for the cycle in which injection occurs.

For detailed information on pulse digital computers, the technical reports of the University of Pennsylvania should be consulted.[1] Pulse counting is also discussed in Vol. 19, Chap. 16.

3·9. Addition and Subtraction by Simple Vacuum-tube Circuits (*E* **or** *I*:*E* **or** *I*).—Vacuum-tube circuits for addition and subtraction may be used where their use is dictated by the requirements of bandwidth, isolation, and impedance level. These methods have been discussed in considerable detail in Vol. 19, Chap. 18, and Vol. 18, Chap. 10. Information here presented is limited to a summary of available circuits and some of their major characteristics; the reader is referred to the previous volumes for details and practical circuits. Vacuum-tube circuits for addition and subtraction are classified as *multiple input*, illustrated in Fig. 3·8; *common plate load*, illustrated in Fig. 3·9; *common cathode load*, illustrated in Fig. 3·10; and the *differential amplifier*, Fig. 3·11.

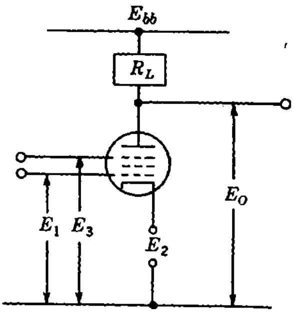


FIG. 3·8.—Multiple input tube adding circuit.

The multiple input circuit is limited by the number of electrodes, introduces weighting factors depending on the tube characteristics, has an input impedance of $1/g_m$ at the cathode and high at the grid, has a tube output impedance of approximately r_p, and can be used for subtraction as well as addition if cathode input is used. Both the common plate and common cathode methods may operate with any number of inputs, each input requiring one tube section; the weighting factors are adjustable if series resistors are used; the input impedances are mainly those due to interelectrode capacitance and grid currents; the output impedance is approximately r_p/n for the common plate circuits and $1/ng_m$ for the common cathode circuits; and neither can be used for subtraction. The differential amplifier usually has two grid inputs,

[1] Technical reports on pulse digital computers include the following published by the Moore School, University of Pennsylvania: J. Von Neumann, "First Draft of a Report on the EDVAC," "ENIAC Progress Report to December, 1943," "ENIAC Progress Report Jan. 1 to June 30, 1944," "Description of the ENIAC and Comments on Electronic Digital Computing Machines," and "Automatic High Speed Computing, a Progress Report on EDVAC." Other reports are "A High Speed Digital Computer," published by the Eastman Kodak Company, Rochester, N.Y., and A. M. Turing, "The Proposed Electronic Calculator," National Physical Laboratory. Previous footnotes have listed popular articles on the subject. Final reports on the ENIAC and EDVAC are being prepared.

although a cathode input can also be used; the weighting factors may be adjusted by resistors; the input impedance at each grid is high; and the input impedance at the cathode is $1/2g_m$. The differential amplifier may be used for subtraction or addition, the voltage between the plates being proportional to the difference between the grid voltages, and the cathode voltage being proportional to the average of the two grid voltages. In all these circuits, the output impedances given must be combined with

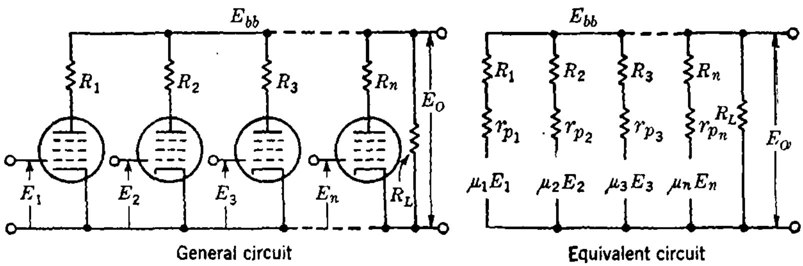


FIG. 3·9.—Common plate-load addition.

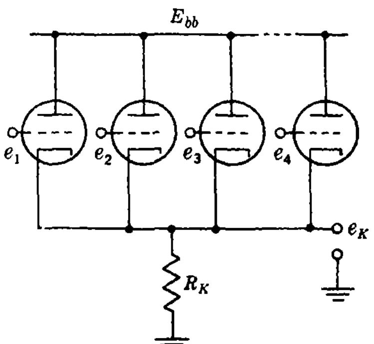


FIG. 3·10.—Common cathode-load addition.

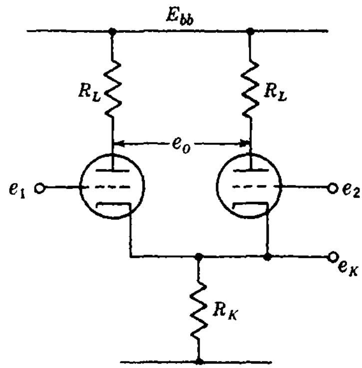


FIG. 3·11.—Differential amplifier for adding and subtracting.

load impedances in the usual manner to obtain the output impedance of tubes and load taken together.

3·10. Discrimination.—Discrimination differs from subtraction in that in subtraction an output is obtained that is accurately proportional to the difference between two input quantities whereas in discrimination the output indicates which of the two inputs is larger but need provide only a rough measure of the difference. Thus, for example, if two voltage generators are series-connected with polarities reversed, the resulting voltage not only goes to zero when the two generators have equal voltages but also is accurately proportional to the difference of the two voltages even when their sum is not zero. On the other hand, a vacuum-tube modulator, because of the curvature of its characteristics, may give a nonlinear output as a function of the difference between two input signals yet accurately indicate the equality of the input signals by a zero signal. The distinction between discriminators and subtraction devices is thus seen to be largely one of quantitative accuracy of output and usage rather than any fundamental difference. Nevertheless, the distinction has been found to be a helpful one.

Discriminators, because of the properties that are stated and implied in their definition, find one of their biggest fields of application in feedback circuits and feedback loops, where they are used to derive error signals representing the inequality of inputs.

Although there seems to be no basic reason why there should not be devices that act as discriminators with the two inputs expressed as different representations, it is nevertheless true that most of the discriminators of importance in the design of computers and other electronic circuits operate on two inputs for which the representation of data is the same. In reading the following descriptions of discriminators, the reader should keep in mind the voids that exist in the field of mixed input representation discriminators. Discriminators with output representation differing from input representation are not at all uncommon.

For certain special applications, discriminators may be made time-sensitive. Thus, for example, a differential amplifier may be used to compare an a-c voltage and a d-c voltage by allowing the tube to conduct only at a time corresponding to the portion of the a-c waveform that is to be compared with the d-c voltage. This might be done by application of a plate voltage pulse or by unclamping the cathode from a high positive voltage. Discriminators may be classified as unidirectional or bidirectional, depending upon whether or not the discrimination action is approximately the same on both sides of the region where the two inputs are equal.

It follows from what has been said that all of the devices for subtraction which have been discussed in preceding sections may also be used for discrimination, although by definition the reverse is not true if accuracy is desired.

Voltage Discriminators (E,E:EorI).—Devices discussed in the category of voltage discriminators will be those in which the inputs are both voltages and the outputs are voltages or currents.

In the *switch-type modulator,* as represented by such devices as the Brown instrument vibrator, two voltages whose variation is slow compared with the vibration frequency (in the case of the Brown vibrator, 60 cps) are discriminated to give an output that is an a-c voltage whose amplitude is indicative of the difference in voltage between input (and may be accurately proportional) and whose phase with respect to the voltage exciting the vibrator indicates the sense of the input inequality. These devices are discussed in Chap. 9, Vol. 19. Equivalent in function are a variety of bridge- and other type modulators using diodes and contact rectifiers, also discussed in that section.

There are available a variety of triode and multigrid tube circuits designed as balanced modulators which are useful as voltage discriminators, the inputs again being slowly varying voltages and the outputs being a-c voltages or currents whose amplitudes are indicative of or proportional to the difference between the inputs and whose output phases relative to the carrier alternating current supplied to the circuits indicate the senses of the input inequalities. A number of such circuits are given in Chap. 9, Vol. 19, and many more may be found in the literature of communication engineering.[1] It is to be noted that these balanced modulation circuits, although having the useful property of gain and high-impedance inputs, are no better with respect to stability than direct-coupled amplifiers.

An ingenious and accurate modulator which may be used either for subtraction or for amplitude discrimination of two d-c (or low-frequency a-c) signals is based on condenser microphone principles. This type of device, developed by RCA for some of their wartime computers and by others[2], consists of a diaphragm or reed put into mechanical vibration by a-c excitation of an electromagnet. Small pickup plates mounted parallel to the diaphragm (or reed) and close to it form small capacitors whose capacitances are varied by the membrane's vibration. If a d-c potential exists across such a capacitor, then the periodic variation of its capacitance will cause an a-c voltage to appear across a resistor in series with the capacitor. The amplitude of this a-c voltage is proportional to the difference in d-c potential across the capacitor, and its phase relative to the a-c excitation reverses as the sign of the difference in potential changes. This device is also useful for changing the data representation from direct to alternating current. A more complete description of this type of modulator will be found in Vol. 19. A photograph will be found in Chap. 19 of the present volume.

Current Discriminators (*I,I:I or E*).—An interesting and occasionally useful device for amplitude discrimination of d-c or low-frequency a-c

[1] See, for example, F. E. Terman, *Radio Engineers' Handbook*, Sec. 7, McGraw-Hill, New York; and F. A. Petraglia, *Electronic Engineering Master Index*, Electronic Research Publishing Co., 1945, pp. 55, 207.

[2] S. A. Scherbatskoy, T. H. Gilmartin, and G. Swift, "Capacitative Commutator," *Rev. Sci. Inst.*, **18**, 415–421, 1947; H. Palovsky *et al.*, *Rev. Sci. Inst.*, **18**, 298–314, 1947; U.S. Patents 2,349,225 and 2,361,389.

currents is the saturable core transformer.[1] It is well known that the effect of saturation of the core of a transformer is the introduction of only odd harmonics of the input into the output unless a constant magnetic bias is present, as by means of a direct current flowing through an extra winding. If such a bias is present, the symmetry of excitation on positive and negative peaks is removed and even-harmonic terms of amplitude roughly proportional to the direct current flowing appear in the output voltage. The phase of the even-order distortion terms reverses

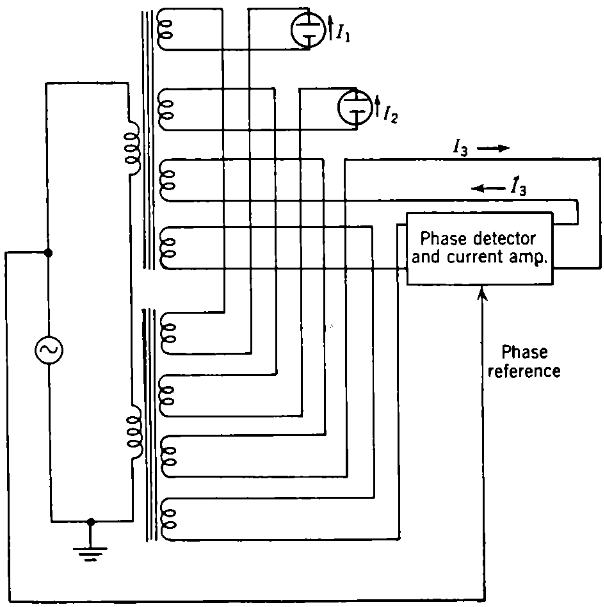


FIG. 3·12.—Saturable core current discriminator circuit.

if the direction of flow of direct current through the third winding is reversed. It is easy to see that if there are several windings carrying direct current on the core of the transformer, the output voltage will contain even-order components unless the algebraic sum of the ampere turns set up by the several currents is zero and that the phase and magnitude of the even-order distortion terms will serve to indicate the direction and amount that any given one of the bias currents should be changed in order to make its magnetizing effect just equal to the algebraic sum of the ampere turns of the other windings.

[1] During the war development work was done on saturable core transformers as current discriminators by the Cornell University Physics Department, under OSRD contract. The work was under the direction of Dr. H. S. Sack. See H. S. Sack *et al.*. NDRC-14 Reports.

Now suppose two such transformers to be connected with primaries and secondaries in series, but with secondaries so phased that their output voltages oppose each other as in Fig. 3·12. With no bias magnetic field, there will be no output voltage from the combination. If magnetic bias of opposite sense is applied to each of the two cores, as by connecting a pair of bias windings in series with the same phasing as that of the output windings and passing a direct current through the pair, the even-order distortion terms in the two secondaries will add, and an output voltage will be observed that may be changed in polarity by reversing the direction of current flow through the bias windings. Pairs of transformers so connected have proved to be accurate and stable current amplitude discriminators. They have further advantages in that the currents being discriminated may be at different voltage levels and that each of the various currents may be multiplied by a scale factor determined by the number of turns of wire in its coil.

As with other discriminators, this device may be used as the error-measuring part of a feedback loop. In this application the output of the magnetic amplifier is amplified and used to control the current in a coil on the magnetic amplifier such as to supply ampere turns nearly equal and in an opposite sense to the algebraic sum of the ampere turns supplied by the other windings.

Frequency Discrimination $(f,f:E)$.—Where frequency is used as a means of data representation, frequency discrimination may be required. An example of such use is the case where the frequency of an a-c voltage derived from an alternator is used as an indication of the velocity of rotation of the shaft on which the alternator is mounted. Where two frequencies are to be compared, the usual practice is to convert to some other representation, with the discrimination in the other representation. Where only a single frequency is to be compared with some fixed standard frequency, a variety of frequency-sensitive bridge circuits and other frequency discriminator circuits are available.[1]

Since phase and frequency are intimately related, phase being the time integral of frequency, phase discrimination circuits may sometimes be used for applications where at first thought frequency discrimination circuits would seem to be required. For example, if it is desired to control a servomechanism such that its average velocity is constant, the frequency of an a-c voltage generated by an alternator on the servo output shaft can be compared with a standard by means of circuits described in the references cited or alternately can be controlled by discrimination

[1] See, for example, F. E. Terman, *Radio Engineers' Handbook*, Sec. 13, Pars. 23–26, McGraw-Hill, New York, 1943; F. A. Petraglia, *Electronic Engineering Master Index*, Electronics Research Publishing Co., 1945, "Frequency Bridges," 175; "Frequency Measurement," pp. 40, 176; "Frequency Meters," pp. 40, 177.

of the phase of the generated a-c voltage and the phase of a standard a-c voltage, which in this case might even be crystal-generated. Phase discrimination circuits are discussed as the next topic in this section.

Phase Discrimination (ϕ,ϕ:E).—Circuits whose output is sensitive to both the amplitude and phase relations of two a-c inputs are generally referred to as phase detectors and have many important applications in the field of servomechanisms and computers and in other fields where sense as well as amplitude must be obtained from a modulated carrier. Two distinct usages of phase information are recognized. In one class of circuits, only phases different by nearly 180 electrical degrees are used, and thus only information as to the sense of an accompanying amplitude modulation is conveyed. A second class of circuits uses phase as a form of data representation. Circuits designed for either type of operation will usually operate properly as phase discriminators, for in this application only rough information as to the sense and amount of phase difference between two quantities is required. Unless special precautions are taken (limiters, etc.), amplitude variation of the input signals will modify the output amplitude, but not the point of phase equality at which the output is zero. Where this phase difference must be measured accurately, the number of available circuits is considerably restricted. For accurate measure of phase difference, phase discriminators may, of course, be used in combination with accurate phase-shifting devices in a feedback arrangement, as has been discussed in connection with other types of discriminators. Phase discriminators have been discussed elsewhere in this series, usually under the title of Phase Detectors; see for example, Vol. 19, Chap. 14, and Sec. 12·12 of this volume.

Some work on phase detectors has been done by the Physics Department of Cornell University, under NDRC sponsorship.[1] There are a number of circuits described in the literature for phase measurement.[2]

Time-interval Discrimination (t,t:E).—Where data are represented by time intervals and two quantities are to be discriminated, a time-interval discriminator is required. This subject is discussed in considerable detail in Vol. 19, Chap. 14, entitled "Time Demodulation," and the reader is referred to this treatment.

[1] See for example, H. S. Sack, and A. A. Olner, "A-c Potential Equalizers and Phase-sensitive Detectors," Cornell Report No. ACE-2, Cornell University, Oct. 26, 1945.

[2] F. E. Terman, *Radio Engineers' Handbook*, McGraw-Hill, New York, 1943, Sec. 13, p. 21; "Phase Indicating Null Indicator for Bridges," *Electronics*, **17**, 242, August 1944; E. T. Ginzton, "Electronic Phase Angle Meter," *Electronics*, **15**, 60, May 1942; "Electronic Phase Bridge for Measurements," *Electronics*, **15**, 96, November 1942; H. Nyquist and S. Brand, "Measurements of Phase Distortion," *Bell System Tech. Jour.*, **9**, 522, July 1930.

Discrimination with other Types of Data Representation.—While discrimination of data expressed in those types of representation most common in electronic computers has been covered in some detail, it is obvious that many types of discriminators for other types of data representation have not been mentioned. It should be remembered that the same concepts apply to other types of data representation, such as linear displacements, shaft rotations, forces, torques, and pressures. Where discrimination is desired between two quantities represented by some physical variable for which the known discrimination devices are not appropriate, discrimination may also be achieved by converting to a new representation and discriminating in the new representation. Conversion from one type of data representation to another is discussed in Sec. 3·21.

MULTIPLICATION AND DIVISION

3·11. Mechanically Controlled Voltage Dividers or Multipliers (E **or** I, S **or** θ:E).—In this section a number of devices will be considered in which the output voltage is proportional to the input voltage multiplied by a constant where the constant is adjustable by mechanically changing the relative orientation or spacing of parts, examples being potentiometers, linear-wound synchros, and condenser voltage dividers. These devices will be considered in this section from the standpoint of linear multiplication, that is, the type of operation in which the transfer constant referred to above is made proportional to some mechanical parameter such as shaft rotation. Such devices modified in their design to produce transfer constants that are nonlinear functions of the input mechanical variable are discussed in Chap. 5.

Potentiometers.—Precision potentiometers are widely used as multiplying devices because of their high accuracy, simplicity, ruggedness, and relatively low cost. They are useful for both d-c and a-f a-c applications. Their practical limitations are those of impedance and speed of response, the latter limitation arising from the fact that they must be mechanically controlled. Potentiometers have been discussed in considerable detail elsewhere in this series,[1] and the reader is referred to these treatments for details.

Precision Variable Autotransformers.—In a manner similar to the potentiometer multiplication described above, multiplication can be performed in the case of a-c potentials by means of a variable autotransformer, sometimes referred to by the trade name Variac.[2] Variable

[1] Vol. 17, Chap. 8; Vol. 19, Chap. 12, entitled "Electromechanical Amplitude Modulation"; and Sec. 12·3.

[2] General Radio Co. trade name. A number of competing devices are also available.

autotransformers on the market at the present time are coarse devices designed for the control of power and do not permit a precision of better than several per cent. Experiments have been made with specially constructed variable autotransformers[1] having particularly homogeneous cores. These were wound very carefully with thin wire to give good resolution and linearity. By carefully adjusting the position of the brush it is possible to obtain a linearity of ± 0.1 per cent with highly reactive loads.

The advantage of the variable autotransformer over the potentiometer is that the linearity of its output voltage does not suffer when a low-impedance reactive load is used. For best results both the load and the transformer should have high Q's. Because of this low-impedance feature, it is possible to cascade several precision autotransformers or to drive other devices such as resolvers with them directly without the necessity for introducing vacuum-tube driver circuits. The inductance of a small autotransformer suitable for computer application is of the order of 2 henries.

Precision variable autotransformers have been used by the British and are designated as "Magslip I-pots." I-pots have been used by the British to drive directly devices such as the 3-in. Magslip resolver. Data on an early British I-pot available to the authors indicates an accuracy of approximately ± 2 per cent, but it is not known whether or not more accurate results have been obtained.

Synchros.—In the usual synchro or synchro-type resolver, the output voltage for a fixed input voltage usually varies with the sine of the angular rotation of the shaft from the zero or null position. However, it is worth noting at this point that synchros have been designed so that the constant relating output voltage to input voltage varies linearly rather than sinusoidally with shaft rotation over a limited range.

Condenser Voltage Dividers.—The subject of condenser voltage dividers in which one of the condensers is controlled by an input shaft rotation is discussed in detail in Vol. 19, Chap. 12.

3·12. Electronically Controlled Voltage Dividers $(E,E{:}E)$.—There appear to be at least two fundamental ways of electronically controlling a voltage divider. In one method an impedance in a divider network is made variable, for example, by varying the plate resistance of a vacuum tube by means of grid control or by varying the power in a high-temperature coefficient resistor operated at a constant ambient. In a second method, an input voltage is switched on and off by a rectangular waveform whose duty cycle is made proportional to another input voltage;

[1] H. S. Sack, J. J. Taylor, and R. N. Work, "Preliminary Results on Calibration of Autotransformer," Report NDRC-364, Cornell University, Research Contract OEMsr768, Jan. 16, 1945.

averaging the resulting output waveform then yields a d-c voltage proportional to the product. The switching in the latter scheme is accomplished by an on-off device such as a vacuum tube alternately driven and cut off, and hence its characteristics are only of second-order importance. A third method, that of varying of the gain of an amplifier, is considered to be distinct from the two methods given and is discussed in the following section.

Variable Impedance Method.—There is little to be said about the variable impedance method except that it is included for completeness but is likely to be extremely inaccurate. It may, however, find uses where only crude but simple multiplication is required, as, for example, in circuits where a stage gain must be changed to maintain a reasonably constant over-all gain but where precision is not important.

Pulsed Attenuator Circuits.—A method has been developed that permits simultaneous multiplication and division and therefore combined operations such as the taking of square roots, squaring, etc., with d-c potentials for inputs and outputs. Referring to the block diagram (Fig. 3·13) (A) designates an attenuator. The input potential A is fed into the attenuator and a fraction of it kA is received at the output. This output kA is fed into an attenuation adjustor, which equalizes the input kA to another potential a_0, by feeding back to the attenuator information that adjusts the attenuation constant k. If now this attenuation constant is applied to a second attenuator (B), identical with the first one, and a potential B is fed into this second attenuator, the output b of the second attenuator will be

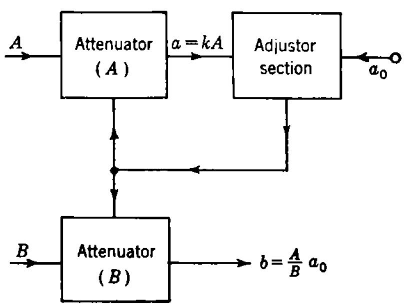


FIG. 3·13.—Pulsed attenuator computer block diagram.

$$b = kB = a_0 \frac{B}{A}, \tag{10}$$

thus performing a multiplication and a division.

Different models differ in the type of attenuators and attenuation adjustors used.

While a simple but rather coarse instrument (accuracy of a few per cent) can be obtained by using an electronically controlled variable impedance as the attenuator, a more precise and more stable device is obtained by designing the attenuator so that the tube characteristics enter only in the second approximation. Such attenuation is accomplished by an electron tube that is intermittently blocked. The transfer constant k is given, in a first approximation, by the ratio of the time during which the tube is blocked to the total time considered. This operation is accomplished by applying to a vacuum-tube grid a rec-

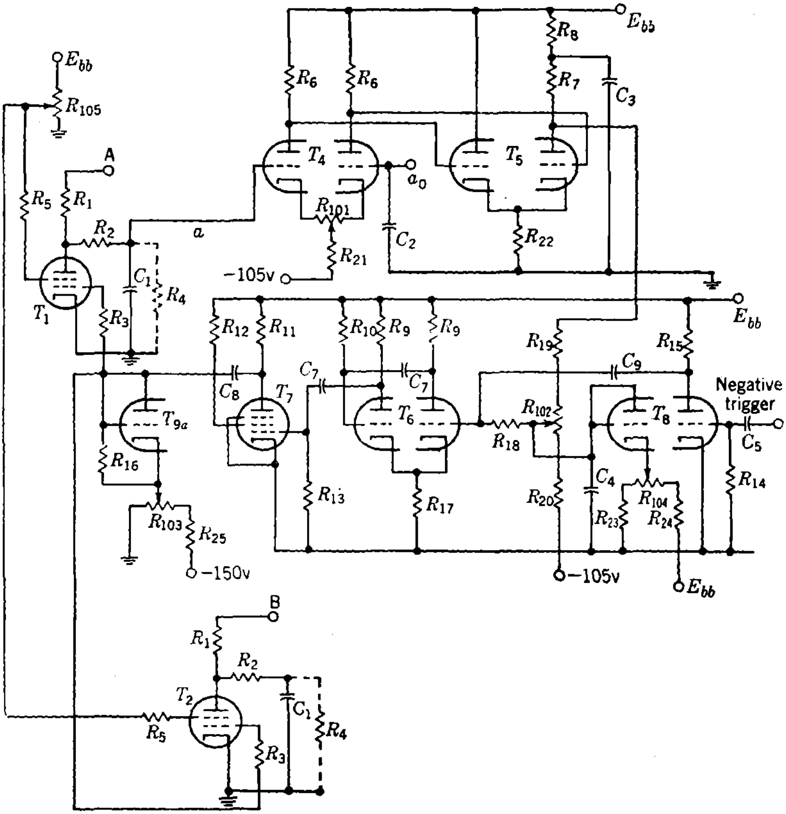


FIG. 3·14.—Pulsed attenuator computer schematic.

tangular waveform in which the relative width of the positive portion is controlled by the attenuation adjustor.

Two different models based on this principle, using different types of attenuators, were developed[1] and will be discussed briefly. One of the

[1] The work on these devices was done for the Radiation Laboratory under OSRD contract by the Physics Department of Cornell University and was under the direction of H. S. Sack. For further details reference is made to NDRC Report No. 14-435 and the Cornell thesis of A. C. Beer and H. W. Boehmer. The material of this section on pulsed attenuator circuits closely follows a summary of this work kindly furnished by Dr. Sack.

circuits is reproduced in Fig. 3·14. The two tubes T_1 and T_2 constitute the two attenuators. The inputs A and B are applied to the plates of these tubes through large resistors R_1. The screen is kept at a constant potential, in this case approximately 45 volts. The rectangular waveform that regulates the attenuation is applied to the grid. In the case of attenuator (A), the output is taken from the point labeled a; condenser C_1 smooths the output. It is evident that while the grid is negative, the condenser will be charged up, whereas it will be discharged during the period when the grid is positive. T_1 and T_2 form, therefore, attenuators of the type required in this kind of circuit. The attenuator output a is fed to one grid of a differential amplifier while to the other grid of this amplifier is fed the other input potential a_0. The differential amplifier consists of two stages of twin triodes T_4 and T_5. Condenser C_3 eliminates oscillations. The output of this amplifier, which is proportional to the difference between a and a_0, is fed through a resistor divider to one grid of a delay multivibrator[1] (DMV), T_6. The DMV is triggered on the grid by a positive trigger source not shown. The output of the DMV is a rectangular waveform, the relative width of which depends on the input d-c potential. This waveform amplified by a pentode T_7 is applied to the grids of the attenuators. The connections are so arranged that if a is larger than a_0, the width of the negative pulse on the grid of the attenuator will decrease, thus decreasing a; if a is smaller than a_0, the negative pulse width will increase, and a will increase. In this way a is always kept nearly equal to a_0 as required. If, furthermore, the resistors and tube clamping impedances in the two attenuators are the same, then the output of the second attenuator is equal to $a_0(B/a)$.

Control R_{101} permits the zero adjustment of the differential amplifier. Control R_{102} regulates the appropriate level of the input to the DMV. A diode, the second half of the twin triode T_8, limits the input potential to values below the critical value at which the DMV oscillates continuously; the correct value is obtained by adjusting the control R_{104}. The adjustments of R_{102} and R_{104} have to be made by observing the output of the DMV on an oscilloscope, adjusting them till the width of the pulse varies over the whole range without changing the mode of operation as the input a varies over its permitted values (or, at a given value of a_0, A varies over its whole range). Another diode, formed by one half of a twin triode T_9, limits the level of the grid of the attenuator tube. This level is adjusted by R_{103} and should be of the order of -3 volts for the tubes that were used here.

The precision that can be obtained by this apparatus depends on a number of factors. The first one has to do with the clamping characteristics of the attenuator tubes. As these tubes are used, the clamped

[1] See Chap. 5, Vol. 20, for details for this circuit.

plate potential is roughly 1 volt or less. Good attenuator tubes would require that the plate resistance in this low-voltage range should be independent of the voltage and should be the same form one tube to another. The 6V6 proved to be the most satisfactory tube in this respect among a great number tested. The input impedance of the following stage, consisting of the leakage resistance of condenser C_1 and the input impedance of the succeeding electronic device, should be high as compared with R_2, constant, and the same for the two attenuators within 1000 megohms, expressed in terms of shunt resistance. Another important factor in the precision is the stability of the differential amplifier, in particular, the zero stability of the first stage.

Tests showed that the precision is somewhat better than ± 0.2 per cent of maximum output with inputs ranging from 23 to 57 volts for a_0, 46 to 230 volts for A, and 22 to 230 volts for B, which means 5 to 165 volts at the output.

A second circuit[1] uses push-pull attenuators. In this circuit, the design of the attenuation adjustors is nearly identical with the first circuit, except that the first stage of the differential amplifier consists of two pentodes instead of a twin triode; the condenser that is inserted to eliminate parasitic oscillations is introduced in a different way; and the various levels have different values. It was found that a precision of ± 0.1 per cent could be obtained when a single-ended input did not exceed ± 45 volts or a differential input did not exceed ± 70 volts. In order to obtain this precision, selected tubes were used.

3·13. Variable-gain Amplifiers and Modulators $(E,E{:}E)$.—Another class of voltage-sensitive transmission constant devices is that employing vacuum-tube amplifiers whose a-c gains are proportional to grid voltages. The 6SK7, for example, has a plate-current grid-voltage curve that is nearly parabolic over the range of grid voltage from -1 to -10 volts, with $e_p = 250$, and $e_{g2} = 100$, as is evident by a substantially linear relationship between g_m and e_{g1} in this region. For such a tube we may write

$$\left.\begin{aligned} i_p &= i_o + ae_{g1} + be_{g1}^2 \\ \text{and} \qquad \Delta i_p &= (a + 2be_{g1})\,\Delta e_{g1} + b(\Delta e_{g1})^2. \end{aligned}\right\} \tag{11}$$

As is apparent from this expression, the gain for small signals of an amplifying stage incorporating a 6SK7 used under these conditions may be varied nearly linearly by adjustment of the d-c bias applied to the grid, thus making possible the multiplication of an input a-c signal applied to the grid of the tube by the magnitude of the increment in d-c bias voltage also applied to the grid. As with the previous method,

[1] See Sec. 6·4 for an application of this circuit.

implying variation of the plate resistance of the vacuum tube, however, this method is subject to considerable instability due to shift in the operating point of the tube with changes in heater voltage.

There is another method employing a variable-gain tube that is somewhat better with respect to variation with changing heater voltage. If in the preceding expressions for the change in the output current resulting from a change Δe in the grid voltage, we set

$$\Delta e_{g1} = \Delta e_1 \cos \omega_1 t + \Delta e_2 \cos \omega_2 t,$$

the result is

$$\Delta i_p = (a + 2be_{g1})(\Delta e_1 \cos \omega_1 t + \Delta e_2 \cos \omega_2 t) + b \left\{ \frac{(\Delta e_1)^2}{2} (\cos 2\omega_1 t + 1) \right.$$
$$\left. + \Delta e_1 \Delta e_2 [\cos (\omega_1 + \omega_2)t + \cos (\omega_1 - \omega_2)t] + \frac{(\Delta e_2)^2}{2} (\cos 2\omega_2 t + 1) \right\}. \quad (12)$$

In this expression the amplitudes of the modulation products are directly proportional to the products of the amplitudes of the input voltages and the grid-cathode voltage does not enter these terms explicitly. This means that variations in d-c bias, cathode temperatures, and the like will be of less importance in determining the amplitudes of these modulation products. If the output of such a stage is passed through a filter transmitting only one of the sidebands, say that at angular frequency $(\omega_1 - \omega_2)$, the output observed at this frequency will be proportional to the product of the amplitudes of the input voltages at angular frequency ω_1 and ω_2, respectively. Computing devices utilizing this property of the tube have been built and have operated with a precision of approximately 1 per cent.

There are available a wide range of circuits of the modulator type that operate to give an a-c output that is the product of an input alternating current and an input direct current. Much is available in the literature on modulators,[1] and the subject has also been discussed elsewhere in this series.[2]

Feedback techniques may be applied to a variable-gain amplifier to cause it to amplify an a-c signal of some standard amplitude such that this signal appears in the output at an amplitude equal to another signal e_1, which is considered as a multiplier signal. This may be accomplished by using the difference in amplitude between the amplified standard signal and the multiplier signal e_1, feeding back the difference at high gain to control the amplification of the variable amplification stage in such a direction as to reduce the difference to nearly zero. Under

[1] For bibliography see F. A. Petraglia, *Electronic Engineering Master Index*, Electronics Research Publishing Co., 1945, "Modulation," pp. 55, 207.

[2] Vol. 19, Chap. 11, entitled "Electrical Amplitude Modulation."

these conditions, the amplification of the stage is proportional to e_1. Any other signal e_2 fed to the variable-gain amplifier in the same manner as the standard signal will then appear in the output with an amplitude proportional to e_1e_2. It is, of course, necessary that the standard signal be separated from the signal e_2 being amplified. The device described is a multiple-purpose element, since two different types of data (e_2 and the standard signal) are used as inputs; any of the methods mentioned in Sec. 1·2 may be used to separate these signals. In this case, separation is easily achieved by using a-c signals at different frequencies for the standard signal and the input e_2. The block diagram of this device would thus closely resemble that of the pulsed attenuator computer (Fig. 3·13).

A multiplying circuit based on these principles has been developed at Telecommunications Research Establishment and is discussed in Vol. 19, Sec. 19·5. A similar method used in this country to linearize the characteristics of a 6AS6 as a variable-gain amplifier achieved excellent accuracies. It is to be hoped that details will be published in the near future by the laboratory undertaking this development, as the method holds considerable promise.

3·14. Special Nonlinear Methods of Multiplication. *Methods.*—There are available a number of mathematical expressions by means of which products can be computed through the use of other functions, such as squares, differences, logarithms, trigonometric functions, etc., which are sometimes easier to instrument than more direct multiplication methods. Three of the most important of these equations are

$$xy = \tfrac{1}{4}[(x + y)^2 - (x - y)^2], \tag{13}$$

$$\log xy = \log x + \log y, \tag{14}$$

and

$$\left.\begin{aligned} xy &= \tfrac{1}{2}[\cos (a - b) - \cos (a + b)], \\ \text{where}\quad x &= \sin a, \qquad y = \sin b. \end{aligned}\right\} \tag{15}$$

Nonlinear Elements.—Nonlinear elements suitable for use in applications of these methods are discussed in Chap. 5 and in Vol. 19, Chap. 19. A short list will be given, but the reader is referred to these other sections for details and for more complete listings. Mechanical nonlinear elements of importance are gears, cams, cone-cylinder combinations, levers, etc. Electromechanical elements of importance are nonlinear potentiometers, loaded linear potentiometers, nonlinear condensers, synchros and synchro-type resolvers, sine potentiometers, etc. Electronic devices of interest are contact rectifiers, nonlinear resistors, vacuum tubes, etc.

3·15. Bridge Methods(Z's:Z) or (θ's:θ).—Bridge techniques are frequently very useful because they provide simple and accurate methods of obtaining an output that is the product of two inputs divided by a third input. Thus, in the bridge illustrated in Fig. 3·15, the output Z_4 is given at balance by

$$Z_4 = \frac{Z_1 Z_3}{Z_2}.$$

The bridge method, as illustrated in Fig. 3·15, has the disadvantage that some balancing means, such as a servo, is required. A practical consideration in the design of such bridges is the adjustment of the slope of the various inputs. The independence of variations of the supply voltage E offered by the bridge makes this method one of great stability and accuracy.

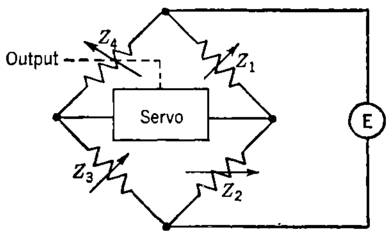


FIG. 3·15.—Use of bridge for multiplication and division.

3·16. Multiplication by the Integration Method.—Bush and Caldwell[1] have reported a method of multiplication that has been found useful in connection with a differential analyzer. In this method the product of two variables x and y is formed according to the equation

$$xy = \int x\,dy + \int y\,dx. \tag{16}$$

Through the use of this equation, multiplication can be performed by standard differential analyzer integrator elements with a very high degree of precision. For discussion of integration techniques, the reader is referred to Chap. 4.

3·17. Miscellaneous Techniques and Devices ($E,E{:}t$) or ($E,t{:}E$)—*Change of Slope of Waveform in Time-modulation Circuits.*—Through changes in the supply voltage, the slope of triangular waveforms used in time-modulation devices may be varied with the sometimes useful and sometimes objectionable result that the time modulation produced by such a circuit changes inversely as the supply voltage when the "pick-off" or comparison voltage is independent of the supply voltage. Expressed in another way, the voltage at any time t after the initiation of the waveform varies directly with the supply voltage. Of course, in the usual time-modulation circuits, the pick-off voltage is also derived from the

[1] V. Bush and S. Caldwell, "A New Type of Differential Analyser," *Jour. Franklin Inst.*, **240**, 255–326, October 1945.

same supply voltage as the triangular (linear sweep) waveform, and this variation is canceled out with a high degree of accuracy.

Use of Variable-condenser Modulators to Multiply A-c Data by D-c Data.—Variable-capacitance modulators are discussed in detail in Vol. 19, are mentioned earlier in this chapter as voltage discriminators, and are also referred to again in Chap. 15. As described above, the output of the variable capacitance modulator is proportional to the amplitude of mechanical vibrations and the applied d-c potential. By the use of an extra electrode, a d-c standard, and a simple feedback circuit, the amplitude of vibration may be made accurately proportional to an a-c input. Under this condition, the resulting a-c output from other electrodes is accurately proportional to the product of the a-c input and the d-c potentials connected to these electrodes. Linearity of modulation may be extremely high, approximately 0.1 per cent. For high-accuracy applications great precision is required in the construction of these devices, and as a result they are difficult and expensive to make.

Field Current Control of Generator Voltage $(\omega,I:E)$ *or* $(\omega,E:E)$.—It has been found useful to vary the volts per rpm scale factor of small a-c or d-c generators by variation of field currents. For example, it may be desirable in a velocity servo to maintain the velocity constant for any given setting of a voltage divider connected to a reference potential. If this is done by comparing with the voltage so derived a voltage from a tachometer generator on the output shaft of the velocity servo, variations in speed will result when the supply voltage changes if the tachometer scale factor is not changed at the same time. It is possible, although not often practical, to control the tachometer in this case by controlling its field in such a manner that the field current is proportional to the supply voltage. The practical difficulties in this case are due to the copper resistivity temperature coefficient and to hysteresis effects.

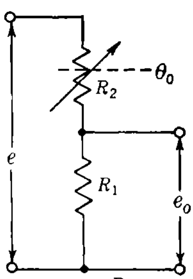
A device that also provides an output voltage accurately proportional to speed and input voltage is the condenser tachometer described in Chap. 4 in connection with integration servos.

Electrodynamometer Multiplier $(I's:I)$.—The electrodynamometer multiplier has been described or proposed by a number of authors.[1] In a typical instrument, two coils are mounted concentrically with one coil free to turn about an axis in the plane of both coils. When current is passed through both coils, a torque tending to produce relative rota-

[1] J. S. Allen, "An Electromechanical Calculator," RL Internal Report No. 62, Jan. 26, 1943; H. S. Sack, Cornell University, "Memorandum on a Computer Based on the Electrodynamometer Principle," July 14, 1945 (unpublished memorandum); V. Bush, F. D. Gage, and H. R. Stewart, "A Continuous Integraph," *Jour. Franklin Inst.*, **203**, 63 (1927); R. N. Varney, "An All Electric Integrator for Solving Differential Equations," *Rev. Sci. Inst.*, **13**, 10 (1942).

tion of the coils is produced, the torque being proportional to the product of the current in each o the two coils, as in an electrodynamic voltmeter. In the electrodynamometer-type multiplier there is provided a second movable coil and a constant magnetic field produced by a coil with fixed current or by a permanent magnet. By passing a current through the second movable coil, a torque that is equal to and oppositely directed from that produced by the first two coils may be produced, which allows operation as a null device. It is seen that the current required in the second movable coil in order to bring the moving coil system to a null is proportional to the product of the currents in the first moving coil and the first fixed coil. A suitable means for detecting small angular motions of the shaft carrying the moving coils must be provided in order to actuate a feedback amplifier supplying current to the second moving coil in order to restore the shaft to its null position. A number of devices suitable for use as shaft position error indicators are proposed in the references, and in addition several of the low-torque data input devices discussed in the first part of Chap. 12 may be used. The instrument, although not at present fully developed, promises to be light, compact, accurate, and practical.

Multiplication and Division Based on Ohm's Law (E,Z:E) or (E,Z:I).—The Ohm's law dividing circuit is a simple and reliable circuit which deserves to be used more often than it is. Its field of application is limited to those cases where the divisor has only one sign and does not go to zero. Because there are so many applications where it is unsuitable owing to these limitations, designers are apt to forget about it for applications where it is entirely suitable. In this circuit, as illustrated in Fig. 3·16, the resistance $R_1 + R_2$, is made proportional to one input y, while the voltage e applied across the two resistors is made proportional to another input x. The current through the resistors is then proportional to x/y and may be measured as a voltage by taking the voltage drop across resistor R_1. The minimum value of y is proportional to the resistor R_1. Since the impedance of this network is variable, the voltage e must be supplied from a low-impedance source or at least measured after R_2 has been adjusted so that $R_1 + R_2$ is proportional to y. This circuit has been used in an airborne navigation computer to compute the time required to fly a given distance at a given speed by making the voltage e proportional to the distance and the resistance $R_1 + R_2$ proportional to the speed, taking the voltage across R_1 as the time required to cover the given distance.



$$e_0 = \frac{eR_1}{R_1 + R_2}$$

FIG. 3·16.—Ohm's law divider.

The Time Coincidence Multiplier (*t's:I*).—A clever method of multiplying devised by A. C. Hardy of Massachusetts Institute of Technology is based on the fact that if several events occur each with a random distribution in time, the probability of a simultaneous occurrence of all of them is proportional to the product of their separate probabilities of occurrence at any given instant. A circuit embodying a similar principle uses repetitive rectangular waveforms whose periods have no common divisor. Under these conditions, the various separate events certainly do not occur randomly with time, but the fraction of the time that coincidence between the positive portions of all waveforms may be observed is very nearly the same as that which would be observed were all events truly random, particularly if the time interval over which observations are averaged is long compared with the repetition intervals of the rectangular pulses and if the frequencies have been judiciously chosen. A tube is adjusted so that it will be turned on only during the time when all input waveforms are positive, with the result that the average current drawn by the tube is closely proportional to the product of the input waveform duty cycles, each of which is made proportional to one of the inputs. An application of this method to the solution of simultaneous equations is discussed in Sec. 6·2. For further details of this circuit, reference is made to Vol. 19, Sec. 19·5.

Computers using this principle have been built with accuracies of the order of 1 to 4 per cent. Since the method does not depend critically upon the characteristics of vacuum tubes or nonlinear elements, there appears to be no fundamental reason why this accuracy could not be improved if necessary.

Miscellaneous Mechanical Methods (*S or θ:S or θ*).—A variety of purely mechanical methods of multiplication are available. Fry[1] has summarized most of these methods. Other references of interest are a U. S. Navy Department pamphlet,[2] a book by Lipka,[3] and Vol. 27 of this series.

The most common mechanical multipliers include logarithmic cams, gears, and tape wheels; mechanical models of two similar triangles, relative sizes being adjusted by one input variable and length of one side being adjusted by the other input variable; "sector multipliers" in which a radius and angle are each adjusted by an input variable, giving a motion proportional to their product over a limited range; the so called "Varigear" positive-drive variable-gear ratio device; space cams; linkage

[1] M. Fry, "Designing Computing Mechanisms," *Machine Design,* August 1945 through February 1946.

[2] "Basic Fire Control Mechanisms," *Bur. Ordnance Pamphlet* 1140, U.S. Navy Dept., Ford Instrument Co., and Arma Corp.

[3] J. Lipka, *Graphical and Mechanical Computation,* Wiley, New York, 1918.

multipliers, etc. For further details and more methods the references given above should be consulted.

IDENTITY OPERATIONS

By an identity operation is meant an operation that does not change the mathematical quantity represented. The principal examples are *change of level, change of impedance, change of scale factor, and change of data representation.* Identity operations are most commonly used in connecting together blocks performing more complicated operations

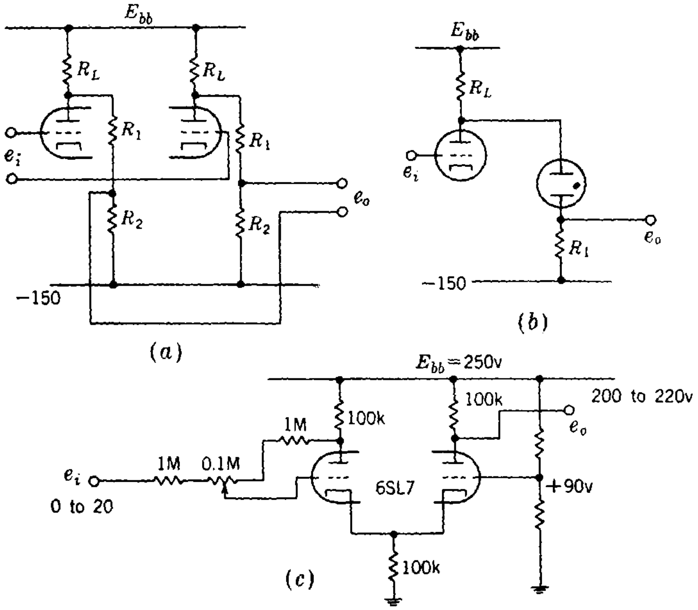


FIG. 3·17.—Examples of circuits for changing d-c voltage levels.

such as multiplication, addition, and squaring. The identity operation of transformation of data representation, for example, allows a block operating on one form of data to be connected to a block operating on a different form of data. In general, it may be said that through the use of identity operations, greater over-all computer effectiveness can be achieved, for the increase in effectiveness of the various other operations will usually more than make up for the complexity of the added identity operation.

3·18. Change of Voltage Level.—Most of the techniques required for changes of voltage level will be found under the discussions of addition and subtraction devices. Circuits shown in Fig. 3·17*a*, *b*, and *c* are examples of level-changing circuits. Figure 3·17a illustrates a method commonly used in direct-coupled amplifiers to obta n signals at the correct level with respect to ground for use as inputs to the following stage. The differential voltage output is only $R_2/(R_1 + R_2)$ times that available at the plate, but a net over-all stage gain results, and both input and output are at the same level. Figure 3·17b, a regulator tube or neon glow tube, characterized by a nearly constant voltage drop over a range of current, is used to obtain a voltage e_0 at a level that is suitable for use as an input for the following stage. A very simple level-changing device is a dry battery. Batteries are bulky, must be replaced frequently, generally cannot be used at very high temperatures satisfactorily, and have large voltage temperature coefficients but are used freely in laboratories and elsewhere where these limitations are relatively unimportant.

In Fig. 3·17c is shown a circuit in which the output is 200 volts higher than the input, the input covering the range of 0 to 20 volts and the output covering the range 200 to 220 volts. The accuracy of this circuit is approximately 50 mv for the range indicated. The circuit may cover larger anges with reduced accuracy.

3·19. Change of Impedance.—Change of impedance is accomplished through the use of networks and/or the use of feedback. These subjects are discussed in great detail in the existing literature and elsewhere in this series. For standard network methods, reference is made to Terman.[1] For feedback theory and examples of the application of feedback techniques to the problem of changing impedance, reference is made to Vol 18, and a recent book by Bode.[2]

3·20. Change of Scale.—Change of scale means multiplication of the data at any point by a fixed quantity. The methods discussed above under multiplication and division therefore are applicable to this problem; but since freedom to vary both inputs as in multiplication is not required for this application, most of the devices mentioned above are unnecessarily complex for this usage. Where the factor is less than one, simple resistive, capacitative, or inductive divider networks will often suffice. Where any impedance must be maintained, a matching T- or π-network may be used. Where the constant of multiplication is greater than one, amplifiers may be used whose gains are accurate. Transformers may also be useful if the distortions and changes in gain with temperature introduced by the transformers are not objectionable. References for the design of matching networks and feedback amplifiers are given in the preceding section.

[1] F. E. Terman, *Radio Engineers' Handbook*, McGraw-Hill, New York, 1943, Sec. 3, par. 25.

[2] H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945

3·21. Change of Representation.—One of the most important identity operations is the change of representation. An introduction to the subject* is found in Sec. 12·2, together with a short list of some of the more important representations of data used in electronic and electromechanical computers.

It is usually true that a computer can be most effective if freedom is allowed the designer in selecting methods and devices for performing the various isolated and grouped operations on the input data. It is obvious that the freedom of representation contributes greatly to the freedom of selection of the best method and device for any operation. For example, suppose that the inputs to a proposed computer are mechanical shaft rotations. It would be imposing unnecessary restrictions on the designer to say that throughout the computer data must be carried along as shaft rotations. A far simpler computer might be made through the use of a simple device performing the identity operation from the representation of data as shaft rotation to the representation of data as an electrical potential (as by a potentiometer), followed by electrical devices performing the necessary operations on the electrical data.

An exhaustive treatment of the subject of the identity operation would probably include a detailed discussion of the devices and methods available for passing from any data representation to any other. Considering only those 17 representations most used in electronic and electromechanical computers† it is seen that $(17)^2$, or 289, categories of devices would have to be discussed, with more than one device in some if not all categories. Such a compendium is not possible here.

It is suggested, however, that the serious worker in the field start his own collection of methods for changes of representation, beginning the collection with changes between the representations most frequently encountered in the type of computers with which he is concerned. The worker in the hydraulic control field will certainly wish to add such variables as volume, flow, etc., to those given above and in all probability can also subtract several from this list as of little practical importance in his field. A discussion of changes of representation useful in passing data into and from an electronic computer will be found in Chap. 12. While the devices of Chap. 12 are treated from the standpoint of data input and output devices for servomechanisms, in most cases the devices are suitable for use also in electronic computers, where servomechanisms may or may not be involved.

* See also H. Ziebolz, *Relay Devices and Their Applications to the Solution of Mathematical Equations*, Askania Regulator Co., Chicago, 1940.

† Force; pressure; torque; translational and rotational displacement, velocity, and acceleration; voltage, current, charge, impedance, frequency, phase, count, and time interval.

Two comments on the use of a change of representation table may be of interest. If the desired device for going from a first given representation to a second given representation is not available or is unsuitable, the change may be made by going to a third representation and then from the third to the second given representation. This is usually impractical, but there are occasional situations where the extra step is justified. A second alternative also exists when the desired device converting from a first given representation to a second given representation is not available or is unsuitable. This alternative is the use of a device converting from the second representation to the first representation, with a feedback loop controlling the data in the second representation on the basis of an "error" signal obtained in the first representation. This will be found to be a useful method for those cases where direct conversion methods do not apply.

CHAPTER 4

CALCULUS

BY J. W. GRAY

4·1. Introduction.—In electronic instruments the need occasionally arises for the differentiation or integration of some quantity with respect to an independent variable. Time is usually the independent variable, and this case will receive most of the attention here. For both differentiation and integration with respect to time, the most common methods used are those employing either a condenser-resistor combination or an electromechanical tachometer. The choice between these two classes in a given application is usually indicated by the nature of the inputs and outputs, but the choice of specific components and circuits requires considerable knowledge and judgment.

Although it is not the purpose of this chapter to embrace all possible forms of the operation of calculus, a few simple methods will be described in addition to time-differentiation integration, including integration and differentiation with respect to a dimension other than time.

DIFFERENTIATION

4·2. RC-circuits. *Simple RC-differentiators.*—The time derivative of a voltage can be obtained directly as a current if a condenser is charged from the voltage source:

$$i = C\frac{de}{dt}. \tag{1}$$

With ordinary sizes of condensers, the current obtained is so small as to be useless as a terminal output except when very rapid voltage changes are measured. The current must generally be converted to the form of a voltage by the use of a resistance. In Fig. 4·1 if the input e_s is changing at a uniform rate and has been doing so for a time sufficient for the current through C and R to become constant, it follows from Eq. (1) that the output voltage is

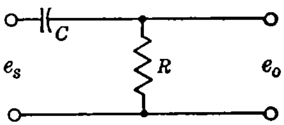


FIG. 4·1.—Simple *RC*-differentiator.

$$e_o = RC\frac{de_s}{dt}. \tag{2}$$

Thus, the larger the RC product the greater will be the output for a given input rate. But also, the larger RC is the more noticeable will be the effects of higher derivatives of the input. For example, if de_s/dt has been constant at some value and then suddenly changes to some new constant value, e_o does not jump immediately to the correct new value of $RC(de_s/dt)$ but rather approaches it exponentially with a decay time constant of RC. In Fig. 4·2, de_s/dt suddenly increases from zero to a constant value (the level of e_s itself is, of course, immaterial), and thereafter the output follows the curve

$$e_o = (1 - e^{-\frac{t}{RC}})RC\,\frac{de_s}{dt}. \tag{3}$$

Thus if RC were doubled, the eventual output would be doubled, but so, also, would the time required to reach this asymptote.

For correct operation, then, RC must be small enough that considering the nature of e_s, the error in the output following any change of the input rate will be of negligible amount and duration. The obvious solution

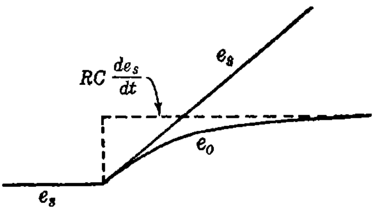


FIG. 4·2.—Effect of transient on simple RC differentiation.

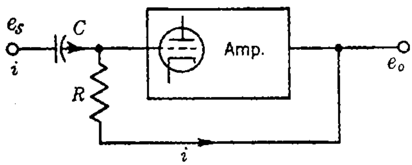


FIG. 4·3.—Feedback amplifier type of differentiator.

is to use a small RC and then to amplify the output to sufficient proportions for the application. There are instances where this method is practical. The amplifier in such a method must usually be of the modulator type rather than direct-coupled, because of the amount of drift encountered in the latter type. With direct-coupled circuits, this drift is amplified. In the feedback method described below, it is not.

Feedback Amplifier RC-differentiator.—The undesirable transient effects in Fig. 4·2 stem from the fact that not all of the signal appears across the condenser; it is apportioned between this and the resistor. This would not be so, if the junction point between R and C could be held at constant potential by a monitoring device that would extract no current therefrom. This monitoring device can be a d-c amplifier. In Fig. 4·3 the amplifier shown inverts (i.e., e_o drops in potential if e_g rises), takes no input current, and has an operating output range that encompasses the input level.

If it is assumed that the amplifier ha infinite gain, so that e_2 must be constant as long as e_o is within its us .l range, the condenser current will be

$$i = C\,\frac{de_s}{dt}, \tag{4}$$

and the output is therefore

$$\begin{aligned} e_o &= e_g - Ri \\ &= e_g - RC\frac{de_s}{dt}, \end{aligned} \tag{5}$$

or, if e_g is at the reference level,

$$e_o = -RC\frac{de_s}{dt}. \tag{6}$$

If the input rate changes, e_g tends to change, which causes e_o to move immediately to counteract this. Thus there is no lag, and the operation is described by Eq. (6) regardless of higher input derivatives.

If the gain of the amplifier is some finite amount G; it is possible to show that the differentiator responds to the type of change shown in Fig. 4·2 as follows:

$$e_o = -\frac{GRC}{G+1}\left(1 - e^{(G+1)t/RC}\right)\frac{de_s}{dt}. \tag{7}$$

Comparing this with Eq. (3), the asymptote, except for the minus sign, is almost the same (only $1/G$ less), and yet the time constant of approach to the asymptote is only $RC/(G + 1)$ instead of RC.

If the input rate is so great that the amplifier saturates before e_o can reach the required amount, e_g will no longer be held in place by the output action, and the device may thus be incapacitated for an appreciable time. For example, if a true step function is fed in ($de_s/dt = \infty$), the grid will follow the step. The amplifier output will jump to the end of its range, and the grid will then move back at a rate equal to this value divided by RC. When it has come back an amount equal to the step, normal operation will be resumed. If the step is positive, grid current may be drawn, which would help charge the condenser and limit the movement at the grid. It is also possible to supplement this effect in the negative direction by the use of a diode. Another method of rapidly restoring the grid voltage employs a relay, which is actuated by the output and operates to shunt R with a low resistance when the output becomes excessive in either direction.

The use of a high-gain amplifier in Fig. 4·3 may result in oscillation. The effect may be analyzed according to ordinary feedback theory by considering the input terminal fixed. The condition may be remedied by the insertion of a small resistor in series with the condenser or by shunting the resistor with a small condenser or both. This comprises a "phase-lead" circuit, which permits a fluctuation of e_o to act on e_g immediately to cancel itself, rather than in a retarded fashion with R and C acting as a filter. If a series resistor of value nR is used or a shunt capacitor of value nC, the effect is the same. Assuming an infinite-gain amplifier, the output is now somewhat imperfect, having a response to the sudden application of input rate of

$$e_o = -(1 - e^{t/nRC})\, RC \frac{de_s}{dt}. \tag{8}$$

This is much the same effect as finiteness of gain, n being comparable in its effects to the factor $1/(G + 1)$.

In some applications it may be that the RC needed to obtain the required output voltage is so large as to be impractical of realization. The output may be increased by using only a fraction of it in the feedback. If the amplifier of Fig. 4·4 is assumed to have infinite gain, the output is

$$e_o = \frac{R_1 + R_2}{R_1} RC \frac{de_s}{dt}. \tag{9}$$

If R_1 and R_2 are not negligible in comparison with R, their parallel resistance simply adds to R.

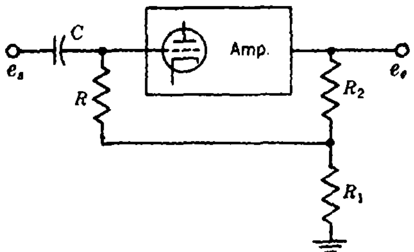


FIG. 4·4.—Increase of output by attenuation of feedback.

The zero drift of the circuit of Fig. 4·3 is just the drift of the amplifier as measured at its input, which is generally quite small in comparison with the available output.[1]

Care is needed in the design of the first stage of the amplifier to ensure sufficient plate voltage that there is no danger of positive grid current.[2] Even so, there will be a small amount of negative grid current (positive-ion flow). By proper choice of the tube type this may be kept fairly low; and if tube selection is permissible, it can be much lower. If this current is constant throughout the operating range, it does not impair the precision, since it produces only a small constant-voltage drop across R, which can be accounted for by the zero adjustment in the amplifier. This may be the case if the amplifier has much gain after the first tube, since this tube will then suffer little change of plate voltage and current.

The e_o point in the circuit is also susceptible to other leakage currents, particularly neighboring high potentials such as the plate lead of the tube. If moisture is a danger, it may be advisable to shield this point

[1] Vol. 18.
[2] *Ibid.*

completely and to connect the shield to a low impedance at the same potential as e_g.

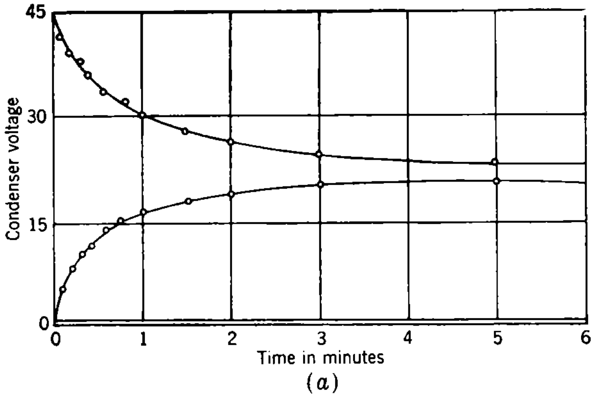
The amplifier output can be at any impedance, depending on the expected external load; for example, an output cathode follower might be used, its nonlinearity and drift being of no consequence if the feedback through R is from its output. The voltage range of the feedback point must extend below e_g by an amount depending on the maximum positive input rate. This is an unusual and troublesome specification for a direct-coupled amplifier. The output has the opposite sign from the input rate. If this is undesirable, a differential amplifier may be used as the output stage, with the output from one plate and the feedback from the other, or a simple paraphase triode may be employed, with appropriate plate and cathode resistors.[1]

Since the operation of the differentiator depends on the RC product, these components must be selected so that this product will be constant, within allowable precision limits, throughout variations of operating conditions such as temperature. One must also take account of condenser leakage, which depends on $e_s - e_g$.

Another important consideration in the choice of condenser is the fact that most condensers are more or less susceptible to the "soaking" or "absorption" phenomenon, which causes them to depart from behavior as ideal condensers. It concerns the inability of a condenser to accept or deliver its entire charge immediately and seems to result from nonhomogeneity of the dielectric. In any event, homogeneous condensers (e.g., polystyrene) exhibit very little of the effect.[2] A simple test is to charge a condenser at a certain voltage for several minutes, discharge it momentarily, and record its subsequent voltage (measured intermittently with a cathode follower so as not to cause discharge). It will rise more or less exponentially to some fraction of the initial charge before decaying downward at its leakage rate. Conversely, if it is solidly discharged for several minutes and connected momentarily to the voltage source, its voltage will drop rapidly for a while before decaying at the normal rate. Figure 4·5*a* shows these voltage-time curves for a typical 4 μf 600-volt filter condenser. The effect amounts to over 40 per cent in magnitude with a time constant of about 30 sec. This aberration can be approximated for purposes of calculation by a combination as in Fig. 4·5*b*. In the case of Fig. 4·5*a*, C' would be almost equal to C, causing the slow redistribution of the momentary charge or discharge as shown by the curves. The value of R' is such that the time constant of the loop is about 30 sec. Such a condenser would obviously be inappropriate

[1] *Ibid.*

[2] J. R. Weeks, "Development of Polystyrene Condensers," *Elec. Mfg.*, April 1946, p. 146.



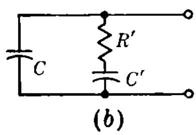


FIG. 4·5.—(a) "Soaking" effect in a representative 4-μf 600-volt filter condenser. The upper curve is for the condenser completely discharged, then charged 1 sec at 45 volts, then open-circuited. The lower curve is for the condenser charged to 45 volts for 10 min, then shorted 1 sec, then disconnected. (b) Equivalent circuit for condenser exhibiting soaking effect.

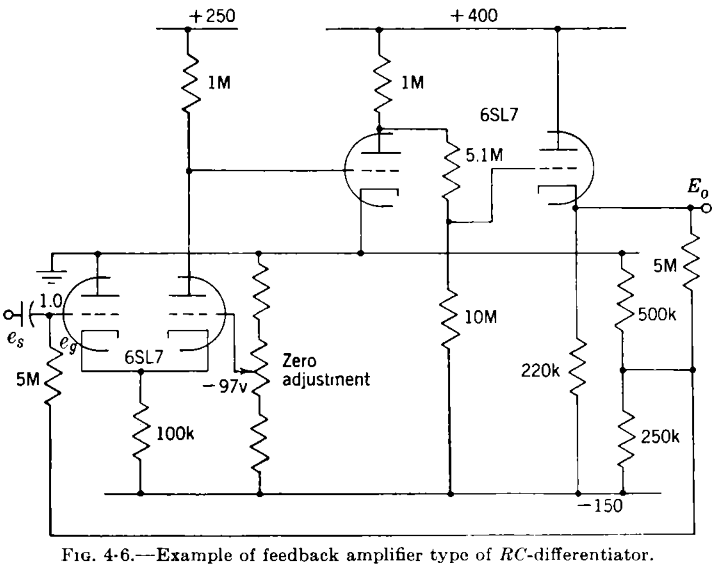


FIG. 4·6.—Example of feedback amplifier type of *RC*-differentiator.

for use in a differentiating circuit, especially if the differentiating resistance R were of the same order of magnitude as the soaking resistance R'.

Figure 4·6 is an example of the type of circuit of Fig. 4·4 where the feedback is attenuated to give greater output for a low input rate. The value of C is 1 μf; R is 5 megohms; R_1 is 5 megohms; and R_2 is the parallel sum of 500 k and 250 k, or 167 k. Equation (9) gives

$$e_o = \frac{(5 + 0.167)(1)(5)}{(0.167)} \frac{de_s}{dt} \tag{10}$$
$$= -155 \frac{de_s}{dt}.$$

The bottom of R_2 is connected, in effect, to -100 volts rather than to ground as in Fig. 4·4. Since $(R_1 + R_2)/R_2$ is 31, this allows e_g to be at about -97 volts, whereas the output e_o is at ground level when $de_s/dt = 0$, a trick that simplifies the amplifier design. The input 6SL7 triode is operated at $\frac{1}{4}$ ma and about 95 volts plate to cathode, ensuring minimum grid current. The differential amplifier input provides a convenient zero adjustment, balances the heater voltage variation effect, and permits the required over-all voltage inversion with two stages. The output is from a cathode follower. It is designed for only positive outputs, as this particular circuit was designed for an application where only negative input rates were of interest.

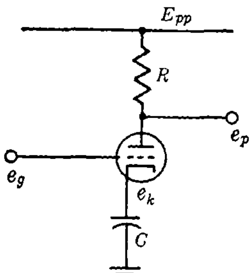


FIG. 4·7.—Cathode-follower differentiator.

Cathode-follower Type RC-differentiator.—For the differentiation of a voltage whose positive rates only are to be measured and whose excursions are somewhat limited in magnitude, an arrangement like that of Fig. 4·7 may be convenient. It has the advantages of simplicity and zero loading of the input. The operation is obvious: If e_g is rising, e_k is forced to rise $\mu/(\mu + 1)$ as fast, and the resulting charging current for the condenser produces an output voltage across the plate resistor. The scale factor relating output to input rate is $\mu RC/(\mu + 1)$. The transient following a change of input rate has a time constant of only $(R + r_p)C/(\mu + 1)$, where r_p is plate resistance.

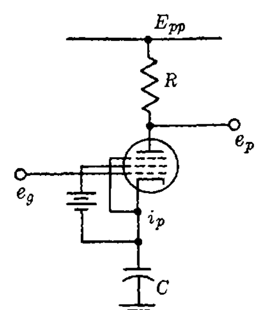


FIG. 4·8.—Cathode-follower differentiator using pentode with separate screen supply.

If a pentode is used, with a separate power supply to hold constant screen-cathode voltage as in Fig. 4·8, the above scale factor becomes RC, and the transient time constant is only C/g_m.

If the screen-cathode potential is maintained constant by means of a d-c "boot strap" device, the screen current will flow in C along with the plate current, so the drop across R will be reduced according to their

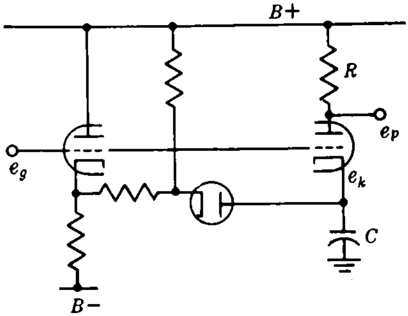


FIG. 4·9.—Method of lowering e_k.

ratio. If the screen is at a fixed potential so the screen-cathode voltage varies with e, the action is more like that of the triode, except for the loss caused by screen current and the fact that plate current is not affected by the drop across R. The fact that the ratio of screen to plate current may vary can make the pentode less accurate than the triode unless the device of Fig. 4·8 is employed.

Grid current is not permissible, and this determines the upper limit for e_g. Heater-cathode leakage can cause a large error, and it may be necessary to use a floating heater supply.

Some method of lowering e_k is needed, since the grid is unable to do it. This may be done with a diode or a diode and cathode follower as in Fig. 4·9. Of course, no useful output is obtained while e_g is descending.

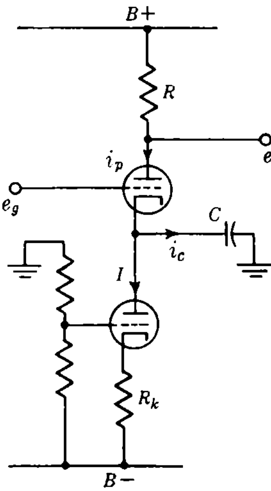


FIG. 4·10.—Cathode-follower differentiator with constant-current tube.

The differentiator can be made to give both negative and positive derivatives by the use of a constant-current device paralleling the condenser as in Fig. 4·10. If I is constant, the output voltage will be reduced by RI. If I varies slightly with e_k, e_p will depend to a small extent on e_g. This may or may not be objectionable, depending on the requirements of a given application.

If the output voltage can be in the form of a potential difference between two terminals rather than a single voltage, the output of the circuit of Fig. 4·10 may be balanced against that of a similar circuit without the condenser to eliminate the dependence of output on input magnitude. This may even permit, in some cases, the use of an ordinary resistor in place of the constant-current device, as shown in Fig. 4·11.

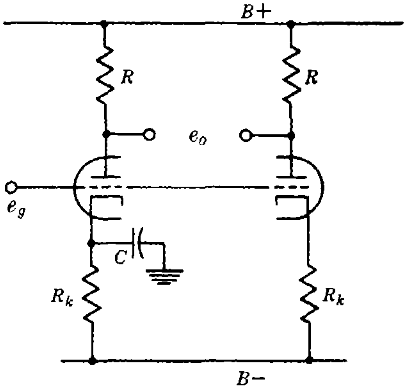


FIG. 4·11.—Balanced cathode-follower differentiator.

4·3. Condenser Circuits Employing Special Current Amplifiers.—There are certain types of feedback amplifiers, employing special devices other than vacuum tubes, that can amplify very small currents to give large currents or voltages with good linearity and stability. One of these employs a pair of magnetic toroids, whose saturation effects develop even harmonics of a carrier voltage when a small d-c unbalance is applied. Another type uses a mirror galvanometer and photocell, with current feedback from a d-c amplifier to cancel out the galvanometer current. Either of these or an equivalent device may be used with a condenser as a differentiator, by amplification of the condenser charging current.

The above magnetic amplifier is described in Sec. 3·10. A differentiator employing an amplifier of this type has been found to be linear and stable to about 1 per cent with input rates ranging from -1 to $+1$ volt/sec. One advantage of this type of differentiator over the type illustrated in Fig. 4·3 is the rapid recovery from the saturating effect of a step function or other temporarily rapid input rate.

A galvanometer photocell amplifier having very good stability and linearity with a full-scale input of 1 μa is described in Volume 18. With a suitable condenser input it would be very satisfactory as a differentiator, although perhaps somewhat limited as to portability.

In these differentiators a resistance must be used to derive a voltage from the condenser current if the output is desired as a voltage. Thus, the RC product determines the relation between rate in and voltage out, and the remarks about this in the preceding section still apply. Also, of course, the condenser should not have excessive leakage or "soaking."

4·4. Differentiation Based on Inductance.—An inductor can be used as a differentiating device if the functions of voltage and current in Eq. (1) are interchanged.

$$e = L\frac{di}{dt}. \tag{11}$$

The RL-circuit corresponding to Fig. 4·1 is shown in Fig. 4·12. The response to a transient input as in Fig. 4·2 is

$$e_o = (1 - Ce^{Rt/L}) \frac{L}{R} \frac{de_s}{dt}. \tag{12}$$

Thus, in contrast to the RC devices, the output may be increased by *decreasing* R.

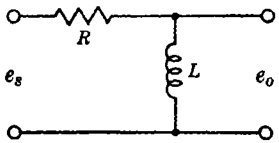


FIG. 4·12.—Simple RL-differentiator.

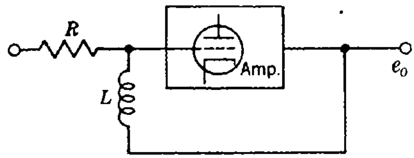


FIG. 4·13.—Feedback amplifier type of RL-differentiator.

All the differentiators described in the preceding section may be made RL devices by replacing R with L and C with R, except that those employing special current amplifiers must use the equivalent voltage amplifiers instead. Figure 4·13 shows the RL equivalent of Fig. 4·3.

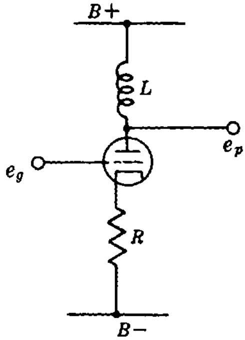


FIG. 4·14.—Cathode-follower RL-differentiator.

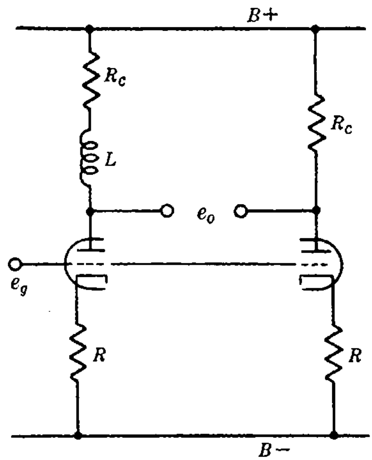


FIG. 4·15.—Compensation of coil resistance.

All the discussion of the latter applies here also, with the above substitution. Figure 4·14 is the RL cathode-follower type of differentiator. This has the advantage over the RC type that no constant-current device is needed to make it operable with negative input rates. When the current is decreasing, e_p can actually rise above $B+$.

One big disadvantage of these circuits in comparison with their RC equivalents is the result of the large amount of resistance in the inductor. This causes the output to shift with the input magnitude according to the ratio of this coil resistance to R. Of course, the same effect obtains in the case of the RC-differentiators, due to condenser leakage, but this can be of a different order of magnitude. If the output may be in the form of a potential difference, a circuit like that of Fig. 4·15 may be employed to cancel this effect. Saturation of an iron-core inductor is perhaps the worst disadvantage. This results in the change with input voltage of the proportionality factor between input rate and output voltage. It is as though in the RC-differentiator, C were a function of the voltage across it.

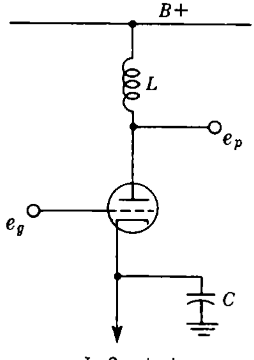


FIG. 4·16.—LC double differentiator.

Hysteresis makes the output characteristic a function of the preceding operation of the device. It is similar to the effects of a "soakable" condenser.

Second-order differentiation may be accomplished by the use of both a condenser and an inductor. For example, a cathode-follower circuit with a condenser at the cathode and an inductor at the plate as in Fig. 4·16 gives an output approximately LC times the second derivative of the input.

4·5. Electrical Tachometers. *Electromagnetic Generators.*—The foregoing types of differentiators are for obtaining the time derivative of a voltage, a quantity that has obvious limitations as to its excursion. When the derivative of a mechanical displacement or rotation is desired and the amount of displacement or rotation is limited, these differentiators could be made applicable by first converting the displacement into a voltage by means of a potentiometer. Often a rotation will have no limits but can continue indefinitely in either direction and with any speed. In this case the speed, i.e., the time derivative of displacement, may be obtained as a voltage by means of a generator. The characteristics of many examples of tachometer generators are given in detail in the *Components Handbook*, Vol. 17, so the discussion here is brief and in general terms.

The most common kind of tachometer generator in electronic instruments is the permanent-magnet d-c type. Suitable small generators are available with proportionality factors ranging up to 10 or 12 volts per thousand rpm. Special features of construction are required in a generator for tachometer use[1] in order to obtain good linearity, stability,

[1] G. Russell, "The Linearity of the Voltage/Speed Characteristics of Small D-c Generators," TRE Report No. T-1999.

absence of brush bounce, and equal output in opposite directions of rotation. Commutator ripple must be minimized for many applications. Commutator wear is troublesome, and commutation at high altitudes and low temperatures presents difficulties. Permanent-magnet material has a high negative temperature coefficient, but the effect of this on the proportionality factor may be greatly reduced by means of magnetic shunts of special materials and by temperature compensation elsewhere in circuits in which the tachometers are used. Current loading must be quite small in comparison with that expected from a power generator of the same size; otherwise linearity will suffer.

Electromagnet fields are used in some tachometer generators instead of permanent-magnet fields. This affords a convenient control of the slope factor, which is sometimes useful. The temperature coefficient of

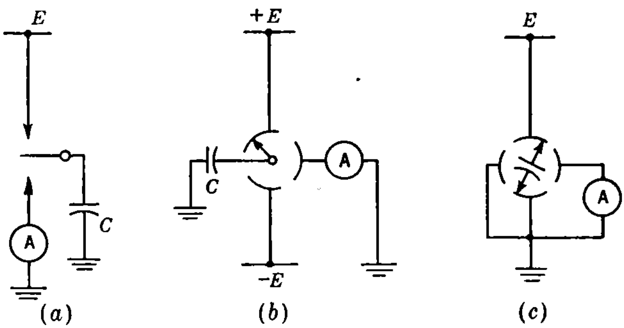


FIG. 4·17.—Condenser tachometer arrangements.

the field is much smaller than that of a permanent magnet as long as the exciting current is constant, but the resistance of any practical field winding has a very high temperature coefficient (over ten times that of a permanent magnet), so a special means must be provided to maintain constant field current.[1] Also, hysteresis of the iron will produce inconsistency if the field current even momentarily rises above the assigned value.

An a-c generator has all the advantages of no commutation, but the generated frequency usually varies along with the voltage, and this may limit the application considerably as to range of speeds. There is a type of induction generator, however, in which the frequency is fixed by an impressed a-c voltage. The rotor is (most commonly) a drag cup, and the carrier and output windings are in space quadrature on the stator. Output voltage is proportional to the product of input carrier and speed, but it is also proportional to the conductivity of the rotor, which makes for a high temperature coefficient. The phase of the output relative to the carrier is affected by speed, and there is some distorted

[1] For a constant-current circuit see Vol. 18.

output at zero speed which is affected by rotor position and is due principally to imperfections of the rotor. The output voltage is much less than that obtainable from the other types.

Commutated Condenser Tachometers.—If a condenser is repeatedly charged to one potential and discharged to another through a current-measuring device, the pulses of charge comprise a current whose d-c component is proportional to the rate of the operation. Thus, if the repeated charging and discharging is done by a switch or commutator on a shaft, the average current is a measure of the time derivative of rotation. In Fig. 4·17*a* the ammeter will give an indication proportional to the speed of switching:

$$i = CEn, \tag{13}$$

where n is the number of switching cycles per second. However, there is no indication of the direction of rotation. Sense of direction may be obtained by the use of more complicated switching sequences or commutation, as illustrated in Fig. 4·17*b* and *c*.

If the current is measured by means of a resistance that develops a voltage, the output will be linear only as long as the voltage output is negligible compared with E. The output curve is actually exponential, leveling off asymptotically as the output approaches E. To obviate this difficulty, a feedback amplifier may be employed as with RC voltage differentiation. The operation of this type of circuit is rendered difficult by the pulse nature of the current; filtering is required, and at very low speeds the output is erratic.

Although these devices are fundamentally differentiators, their common application is as part of an integrator. More will be said about them under that heading.

4·6. Mechanical Differentiators.—This section is limited to a brief mention of some of the types of mechanical differentiators that might be used as auxiliary devices in apparatus that is primarily electronic.

Drag-type Differentiator.—These differentiators, of which a common example is the automobile speedometer, give a mechanical displacement output proportional to the rate of the mechanical input. The output is coupled to the input through a medium such as a viscous fluid or a magnetic eddy-current device so that the force or torque exerted through the coupling is a function of the difference in velocity between input and output. The output is constrained to relatively small excursions by a spring. Ideally, the spring obeys Hooke's law, and the coupling is such that the torque is proportional to the velocity. Under these conditions, the displacement of the output is proportional to the velocity. A metal disk or drag cup revolving in a magnetic field gives a fairly linear torque-velocity characteristic. If the torque is nonlinear, the spring may be designed in such a way as to cancel this nonlinearity, yielding an output displacement linear with respect to input rate.

The response to higher derivatives is similar to that of the simple *RC* voltage differentiator. For example, if the input rate changes abruptly, the output moves exponentially (if the inertia is negligible) to its new value, as in Fig. 4·2, with a time constant equal to the proportionality factor between output displacement and input rate.

Gyroscope.—A spinning mass may be rotated about an axis perpendicular to its axis of spin if it is constrained not to rotate about the axis perpendicular to both of these, but it will exert a torque against the constraint proportional to the speed at which it is being rotated. Thus, if the constraining torque is obtained from a spring that allows a small amount of deflection, the deflection is a measure of speed and may be made available as an output in any of several ways, as in the foregoing paragraphs. This deflection may be several degrees without appreciably affecting linearity, since cos α is very nearly unity when α is small. Spin velocity must be constant, since the torque is also proportional to it. Very slow rotational speeds can be measured quite accurately by this method. A common application of the method is the rate-of-turn meter in aircraft.

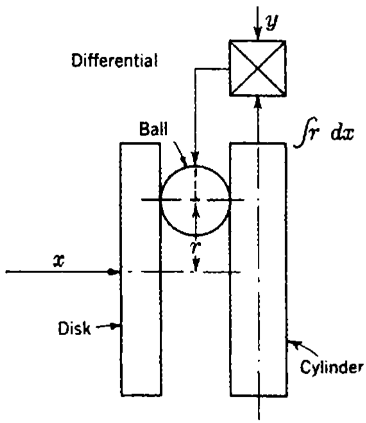


FIG. 4·18.—Ball-and-disk integrator as a differentiator.

Accelerometers.—A second time derivative may be obtained by the use of a spring and a mass.[1] If the mass is accelerated, the accelerating force is proportional to acceleration; and if this force is applied through a spring, a deflection results that is proportional to acceleration. This principle can be applied to measure the second derivative of either translational or rotational motion. Any of several types of pick-off may be used for the output, as above.

Ball-and-disk Differentiator.—The ball-and-disk integrator, which is described under that heading, can be used with feedback to obtain the derivative of one motion with respect to another. In Fig. 4·18, one of the motions x drives the disk, and the other y is compared with the cylinder output by means of a differential, the difference, or error, being used to position the ball. The cylinder output is $(r/a)\ dx$, where r is the position

[1] See, for example, C. S. Draper and W. Wrigley, "An Instrument For Measuring Low Frequency Accelerations in Flight," *Jour. Aero. Sci.*, **7**, 388–401, July 1940.

of the ball and a is a multiplier depending on dimensions. If r is static at its correct value, the output rate is equal to the rate of y, so $dy = (r/a)\,dx$ or $a\,dy/dx = r$, so that the position of the ball is a measure of the derivative of one input with respect to the other. If r is incorrect, the differential will act to correct it. In this case, the equation at the differential is

$$dr = g\left(dy - \frac{r}{a}\,dx\right), \tag{14}$$

where g is the mechanical amplification in positioning the ball. If dy/dx has been zero and changes to some constant amount, r will approach its correct position according to the exponential formula

$$r = (1 - e^{-\frac{gx}{a}})\,a\,\frac{dy}{dx}. \tag{15}$$

The detection of the error and control of r might be done better by electrical or other more elaborate means, so as to increase g and decrease loading of the integrator.

INTEGRATION

4·7. RC Integrating Circuits. *Simple RC Integrator.*—A condenser alone will give the time integral of a current in the form of a voltage

$$e = \frac{1}{C}\int i\,dt. \tag{16}$$

But an input is seldom obtainable directly in the form of a current. A voltage input may be converted to a charging current for the condenser by means of a resistance, as in Fig. 4·19. But since e_s is divided between R and C, if the voltage e_o across C becomes an appreciable part of e_s, i will not be an accurate measure of e_s.

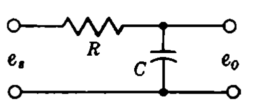


FIG. 4·19.—Simple RC integrating circuit.

If the input rises suddenly from zero to some fixed value, the response of a perfect integrator would be linear, as shown by the dashed line in Fig. 4·20. The actual output of the RC-circuit is the exponential curve shown:

$$e_o = e_s(1 - e^{-\frac{t}{RC}}). \tag{17}$$

Compare this result with that of the differentiating circuit (Fig. 4·2); in the present case the output corresponds to the ideal only at the start of the transient, while the output of the differentiator was in error at first and gave the desired result only after the transient had died out.

Feedback Amplifier RC Integrator.—The charging current may be made to be a true measure of e_s by the use of a d-c amplifier as in Fig. 4·21, where e_g is held constant or nearly so by the inverse feedback through the condenser. As in the case of the similar differentiating circuit, the amplifier must be arranged to invert and must draw no appreciable input current. It is not necessary here that the output voltage range encompass the grid level. On the other hand, the grid level is the reference voltage for e_s, as these points are conductively coupled through R.

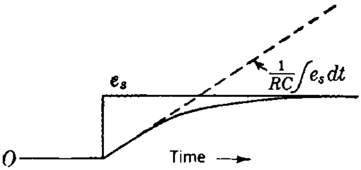


FIG. 4·20.—Response to a step function.

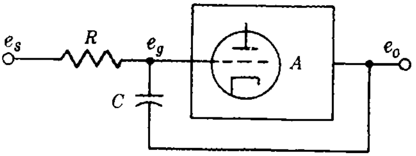


FIG. 4·21.—Feedback amplifier RC integrator.

On the assumption of infinite amplifier gain, so that e_g is not permitted to move at all while e_o is within its operating limits, the condenser current is

$$i = -C\frac{de_o}{dt}. \tag{18}$$

But this is the current which flows in R; so if e_s is measured with respect to e_g,

$$e_s = -RC\frac{de_o}{dt} \tag{19}$$

or

$$e_o = -\frac{1}{RC}\int e_s\,dt + \text{(constant)}. \tag{20}$$

The constant of integration is simply the voltage of e_o at the start of the operation.

Taking the amplifier gain G into account,

$$e_o = -Ge_g \tag{21}$$

and

$$\begin{aligned} i &= C\frac{d(e_g - e_o)}{dt} \\ &= -C\frac{G+1}{G}\frac{de_o}{dt}. \end{aligned} \tag{22}$$

But

$$i = \frac{e_s - e_g}{R} = \frac{e_s}{R} - \frac{e_o}{GR}, \tag{23}$$

so

$$\frac{de_o}{dt} = \frac{-G}{G+1}\frac{e_s}{RC} - \frac{1}{G+1}\frac{e_o}{RC}. \tag{24}$$

For a step function as in Fig. 4·20, this integrates to

$$\begin{aligned} e_o &= -G[1 - e^{\frac{-t}{(G+1)RC}}]e_s. \qquad (25)\\ &= -\frac{G}{G+1}\left[t - \frac{t^2}{2(G+1)RC} + \cdots\right]\frac{e_s}{RC}. \end{aligned}$$

Thus e_o approaches a value of magnitude $-Ge_s$ with a time constant of $(G+1)RC$ (Fig. 4·22) rather than approaching e_s with a time constant of RC, as for the simple integrator. The initial rates of integration are nearly the same [$1/RC$ compared with $G/(G+1)RC$], but the duration of the integration before a given amount of error occurs is increased by a factor of $G+1$. Generally, the gain is such that for ordinary values of e_s the asymptote is far beyond the limits of the output range, so that e_o would be leveled off abruptly by saturation long before approaching $-Ge_s$ as in Fig. 4·22.

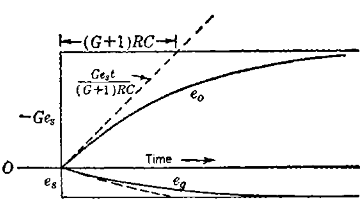


FIG. 4·22.—Response of circuit of Fig. 4·21 to step function.

These integrators are often referred to as "feed-back time constant" circuits and sometimes as "Miller feedback" circuits because the condenser has the same qualitative effect as grid-plate capacitance. It must be emphasized that though the time constant is greatly increased by the amplifier, the output rate for a given input is hardly changed; in fact, if the gain is infinite, it is not changed at all. The grid voltage, though, behaves as would the output of Fig. 4·19 if the condenser were $(G+1)$ times as large. In some filtering applications this point is used as the output rather than e_o, the amplifier being for the sole purpose of multiplying the effectiveness of the condenser.

The fact that the output range need not include the grid voltage, because the feedback is via the condenser, makes for simpler amplifier design than in the case of the differentiator. This is fortunate, as the problem of voltage integration seems to arise more frequently in computers than that of voltage differentiation.

A very common type of integrator employs a single pentode or high-μ triode as in Fig. 4·23. The reference level for the input is the value that e_g has when e_o is at its neutral value. If it is desired that this input reference be ground, the cathode must be raised by the use of a bleeder or by a self-biasing resistor. In the latter case the gain is somewhat reduced—the improvement of linearity afforded by resistive degeneration being of no value in this respect.

The effect of condenser charging current on the relationship between e_o and e_g was ignored in the foregoing equations. If circuit parameters and integration rates are such that it becomes appreciable compared with plate current, the integration will be adversely affected. In particular, if e_s goes so far negative that the condenser takes all of the current in R_p, the tube will cut off and the output will be in error.

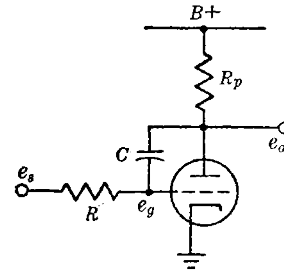


FIG. 4·23.—One-tube integrator.

Accuracy of integration may be evaluated in terms of the constancy of e_g in proportion to the input voltage as e_o traverses its normal range. If e_g changes by Δe_g as e_o moves from one operating limit to the other, the proportional change in integration rate for a given input is $\Delta e_g/(e_s - e_g)$. Thus, the greater the input the less will be the percentage error. On the other hand, if the scale of the input is increased relative to the scale of the output rate, RC must be increased accordingly. For long-time integration it is sometimes necessary to keep the input scale rather low so that R and C will not be excessive. In this case a very high gain amplifier is needed, and amplifer drift will be more noticeable.

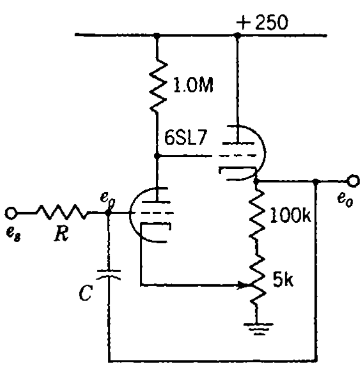


FIG. 4·24.—Integrator with resistive regeneration and cathode-follower output.

The negative condenser feedback gives the amplifier great stability and freedom from oscillations, as compared with the analogous differentiating circuit. This makes it feasible actually to attain perfect integration through the use of resistive regeneration to obtain infinite gain in the amplifier. The circuit of Fig. 4·24 is an example of this and also of the inclusion of a cathode follower in the amplifier to obtain a low-impedance output and to eliminate the effect on the amplifier of the condenser current, though the latter is rarely a serious consideration. The feedback to the cathode of the amplifier tube may be adjusted so that for a limited range of e_o, practically no movement takes place at the grid.

The regeneration can be increased *beyond* the point of "infinite gain," and the device will still be operative as an integrator, with the error in the opposite direction as compared with the finite-gain case. That is,

G will reverse its sign, so that the curvature of e_o in Fig. 4·22 reverses. Figure 4·25 shows the voltage characteristics of the amplifier (with a 5-megohm resistance in series with the grid to show the limit imposed by grid current), with three different settings of the regeneration. Without the RC combination the reflex portion of curve C would not be realizable and the output would follow one or the other of the dashed lines as e_g is raised or lowered. With R and C operating and e_s at some fixed positive or negative voltage, e_o and e_g will slowly trace out all of curve C, with the drop across R, and therefore the output rate, actually increasing during the reflex portion. It is not permissible to obtain the negative gain simply by reversing the sense of the amplifier, for the resulting device would then behave like a trigger circuit. The condenser feedback must be negative.

The above amplifier requires careful adjustment which depends on the μ of the tube and at best gives accurate integration over a very

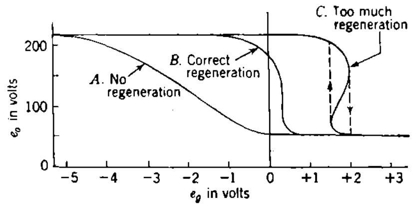


FIG. 4·25.—Characteristics of circuit of Fig. 4·24.

limited output range. Also, there is no convenient means for adjustment of the input reference level, and this will change considerably with change of heater voltage (e_g being lowered about 0.1 volt by a 10 per cent rise of E_f). Reference is made to Vol. 18 for examples of stable amplifiers having larger output ranges with high gain.

All the remarks in Sec. 4·2 about grid current, leakage to the grid terminal, constancy of RC, and condenser soaking apply also to the integrator. Condenser leakage, in so far as it may be represented by a constant shunt resistance, has the same effect as reducing the amplifier gain. It actually comprises a negative feedback of R/R', where R' is the leakage, and thus may be neutralized by positive feedback in the same ratio.

4·8. Integration Based on Inductance.—If a voltage is impressed on an inductance, its time integral appears as the resulting current:

$$i = \frac{1}{L} \int e\,dt. \tag{26}$$

A voltage or power output could be obtained from this by means of one of the special current amplifiers mentioned in Sec. 4·3.

The *RL* integrator analogous to the *RC* device of Fig. 4·19 is shown in Fig. 4·26, where *R* replaces *C* and *L* replaces *R*. The *RC* equations also apply here if L/R is substituted for *RC*.

Similarly, a feedback amplifier type of integrator may be constructed employing inductance, as in Fig. 4·27. A current flows in *L* proportional to $e_o - e_g$, so, unlike the *RC*-integrator, it is advisable to design the

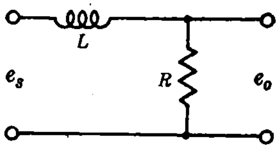


FIG. 4·26.—Simple *RL*-integrator.

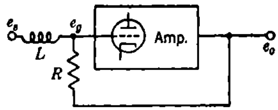


FIG. 4·27.—Feedback amplifier *RL*-integrator.

amplifier with its output range in the neighborhood of the input potential, since this current tends to saturate the inductance.

A cathode-follower arrangement may be employed analogous to the circuit of Figs. 4·7, 4·8, 4·10, or 4·11, an example being shown in Fig. 4·28.

Except for very short time integration, it is not practical to use air-core inductors. All the remarks in Sec. 4·4 concerning the coil resistance, saturation, and hysteresis apply analogously in integration.

4·9. Integrators Employing Tachometers. *Integration of a Voltage with Respect to Time.* Any of the tachometers described in Sec. 4·5 for obtaining the time derivative of a mechanical rotation in the form of a voltage may be used as the basis for an integrator that gives the time integral of a voltage in the form of a rotation. In this case the tachometer is driven by a motor, and their rotation comprises the output. The input voltage is compared with the tachometer output voltage, the difference being fed to a power amplifier whose output drives or controls the motor. The basic circuit is exemplified in Fig. 4·29, although there are other input arrangements than the series circuit shown.

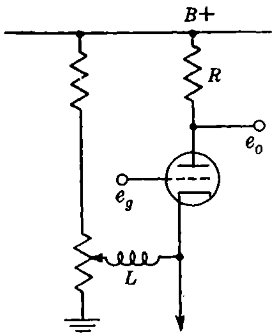


FIG. 4·28.—Cathode-follower *RL* integrator.

The amplifier operates to drive the motor and tachometer at such a speed that the amplifier input e_g is held within very close limits of the reference level. The sense of the amplifier is such that if e_g is excessive one way or another, the motor will accelerate or decelerate so as to bring it back toward zero. Thus the tachometer speed is made to be such that its voltage output is very nearly equal to e_s, so the total shaft rotation is a measure of the time integral of e_s.

This integrator is actually a servomechanism and is usually referred to as a rate servo or velocity servo. The servo aspects of the device are discussed in Secs. 14·4 and 14·5. Several examples are given in detail, with methods of employing some of the various types of tachometers that were described in Sec. 4·5.

The higher the gain of the amplifier the less will be e_g and the more nearly equal tachometer voltage e_t will be to e_s. If the motor speed

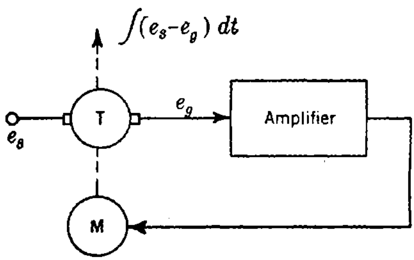


FIG. 4·29.—Velocity servo integrator.

were proportional to e_g, however, e_t would still be proportional to e_s and the integration would still be satisfactory. But mainly because of static friction, this is by no means the case, especially where the motor is required to run in either direction. This is illustrated in Fig. 4·30 which shows a large "dead space" at zero speed, with considerable error signal required to start the motor and generator in either direction. The dashed line is the best linear approximation to the actual curve, and the proportional error of integration will be the difference between this and e_g divided by e_s. Thus, the larger the voltage range covered by e_s and e_t the less will be the resulting percentage of error. Also, of course, the amplifier gain should be as high as possible.

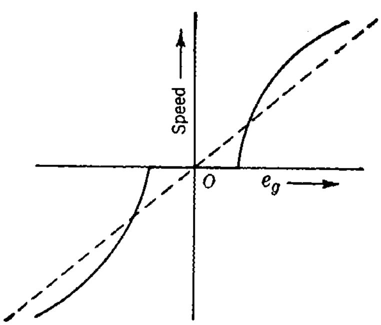


FIG. 4·30.—Typical velocity servo response to error signal.

In contrast to the RC and RL integrators, the rate servo integrator operation is not affected by the magnitude of the integral, i.e., the total rotation of the shaft. It may be affected, however, by fast changes of the input voltage. If this changes abruptly or at a greater rate than it is possible for the motor-generator combination (and associated mechanical parts) to follow because of inertia, part of the integral of the input will be left out of the total output. For example, in Fig. 4·31, assume that e_s has been constant, corresponding to a certain rate of integration.

Assume, also, that the amplifier gain is infinite, so e_t equals e_s. This assumption does not imply that it will never saturate but only that within limits of saturation, no movement is required of e_g. Now e_s is rapidly increased to some new value, at a rate faster than the available acceleration of the mechanical system with full accelerating power. The speed finally comes up to the correct value, as shown by equality between e_t and e_s, but an error in the total rotation has accrued, equal to the integral of $e_s - e_t$; i.e., the integral of the error voltage e_g.

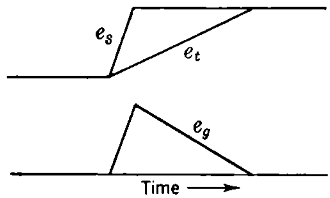


FIG. 4·31.—Illustrating loss of integration caused by rapid acceleration.

There is a way of neutralizing this acceleration error by combining with the amplifier an RC integrator (Sec. 4·7), which evaluates the error $\int e_g\, dt$ and makes the motor run at speeds beyond the required new speed (Fig. 4·31) until this integral is brought back to zero. As shown in Fig. 4·32 two amplifiers are required; a high-gain voltage amplifier and a power amplifier to drive the motor. The latter should have sufficient gain so that very little movement of e_o, relative to its available swing, is required to give full power output in either direction. The first amplifier with its condenser feedback operates to hold e_g constant at the zero level in spite of any deviation of e_t from e_s. If there is an error as in Fig. 4·30, a current will flow in R proportional to the error, which will flow in C by virtue of movement of e_o (see Sec. 4·7), and the resulting

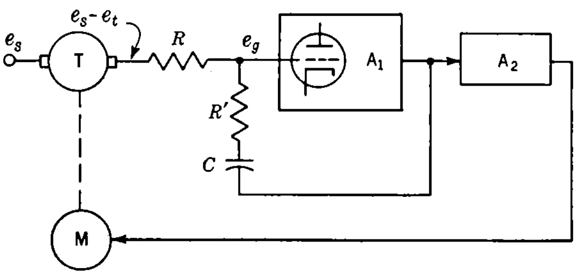


FIG. 4·32.—Use of RC integration to cancel acceleration errors.

charge in C will measure the integral of $e_s - e_t$. Before e_o can get back to its normal value, this charge must be delivered back in the form of reversed current in R, so $e_s - e_t$ will reverse its sign until the error is just canceled. The auxiliary resistiance R is to prevent oscillation, as is explained in Sec. 14·5; its presence does not affect the above cancellation. The first amplifier may be given infinite gain by internal resistive regeneration (Sec. 4·7): in this case the net error signal will be zero.

Of course, the accuracy of the integrator is no better than that of the tachometer employed. These were discussed briefly in Sec. 4·5, and more details are given in the *Components Handbook*, Vol. 17.

A d-c input voltage may be integrated by the use of either a generator or a commutated condenser type of tachometer. In the latter case the feedback required to keep the tachometer from having an exponential characteristic, as mentioned in Sec. 4·5, obtains directly as part of the servo loop. The details are given in Sec. 14·5. An advantage of the condenser over the tachometer generator, in some applications, is that if the input voltage and the condenser-charging voltage are derived from the same source, the integration will be independent of variations in the source.

An a-c input voltage, if it is not required to pass through zero, may be simply peak detected and applied as a d-c voltage. If it does go

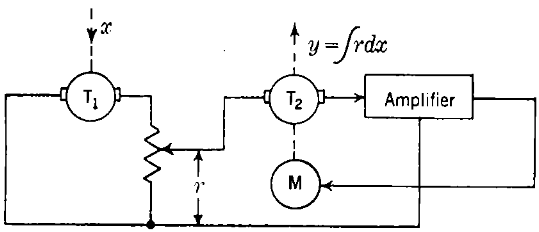


FIG. 4·33.—Integration of one motion with respect to another.

through zero, a reference a-c voltage is needed to discriminate between positive and negative inputs. A phase-sensitive relaying circuit may be employed to switch the direction of the servo output when the input passes through zero, or a linear phase-sensitive detector may be employed to convert it to direct current.

An a-c voltage may also be integrated by means of an induction generator type of tachometer, with the advantage that the error signal is alternating current and may be amplified as such. If the reference alternating current applied to the excitation winding of the generator is from the same source from which the input is derived, the integration will be independent of source variations, as in the case of the condenser tachometer. The primary disadvantages of the induction generator are the great temperature sensitivity, the speed-sensitive phase shift between excitation and output, and the low "signal-to-noise ratio" at low speeds.

Use of Velocity Servos to Integrate with Respect to a Dimension Other Than Time.—In combination with an additional tachometer and a potentiometer as in Fig. 4·33, a velocity servo can produce a rotation that is the integral of one displacement or rotation with respect to another. The dimension of time, although it plays an important role in each part of the device, need not appear in the over-all equation. If x is the total rotation put into the first tachometer T_1 and r is the potentiometer setting, the voltage delivered to the rate servo is proportional to $r\,dx/dt$. But the rate servo output is proportional to the time integral of this, or simply $\int r\,dx$.

Unless the accelerations of x can be kept under the value that would require greater acceleration of T_2 than is available, it is necessary to employ the RC integrator combination or its equivalent in the amplifier. The integration in this case will not be immediate but will eventually reach its correct value.

This integrator performs the same function as the ball-and-disk integrator of Sec. 4·11, in that both of the inputs and the output are mechanical. The outstanding advantage, for certain applications, is that all three motions as well as the amplifier may be mechanically remote. Also, it may be assembled from standard components.

By the use of nonlinear potentiometers, the integration with respect to x of various types of functions may be achieved. Also, several input tachometers and potentiometers may be added in various ways for the input, to obtain combinations of integrals.

4·10. Watt-hour Meters as Integrators.—An ordinary induction-type watt-hour meter used as such gives the time integral of a-c power as a shaft rotation. It may be used to integrate an a-c voltage, impressed on one of the two windings if a reference voltage is impressed on the other.

This latter must be of constant magnitude and phase. In the case of the velocity servo integrator employing either a condenser or induction generator tachometer, if the reference voltage and input vary in proportion, the integration rate remains constant. This is not true of the watt-hour meter; in fact, if the reference and input are derived from the same source, the integration rate will vary as the square of the source voltage.

The principle of operation is that the torque produced on a metal disk by the action of the fields of the two currents, as in an induction motor, is opposed by the eddy-current drag of a permanent magnet field. The first torque is proportional to the product of the two currents, while the second is proportional to the rotational speed; thus they are equal when the speed is proportional to the product of the currents. Both torques are proportional to the conductivity of the disk, which, therefore, cancels in the equation. The permanent magnet will strengthen with decreased temperature, giving slower integration as with the d-c tachometer, but special magnetic shunts may be used to offset this.

The fact that the integration rate is proportional to the cosine of the phase angle between input and reference voltages may be put to advantage in certain types of computers. For example, if it is desired to combine integration with the resolution of a vector into rectangular components, this is accomplished directly if the vector angle corresponds with the above phase angle.

4·11. Mechanical Integrators. *Gyroscope.*—If a torque is applied to a free gyro about an axis perpendicular to its axis of spin, it will not yield to this torque but will, instead, rotate about the axis perpendicular to both the axis of spin and the axis of the torque. The rate of this rotation, or "precession," is proportional to the torque; therefore the total rotation is a measure of its time integral. The spin velocity must be constant, as the integration rate is inversely proportional to it.

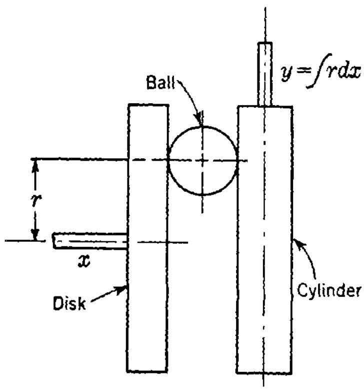


Fig. 4·34.—Ball-and-disk integrator.

The input torque may be derived from a displacement or rotation by means of a spring, but it must be applied in such a way that its axis can rotate so as to remain at right angles to the spin axis whose rotation comprises the output. If this is not done, the integration rate will decrease with the cosine of the output rotation. The input may be applied electrically by means of an induction torque motor. In this case the resistivity of the torque motor rotor will affect the rate. Since no actual rotation is required of the torque device on its own axis, it may be an electrodynamometer or permanent-magnet and coil device as in an ammeter.

An interesting application is found in the German V-2 projectile, where the acceleration is integrated by means of a gyroscope to determine the velocity. The precessing torque is obtained from the acceleration of the inertia of the mass of the gyroscope and is thus proportional to acceleration.[1]

Ball-and-disk Integrator.[2]—This purely mechanical device is sometimes used as a link in electronic computers where it is necessary to integrate a variable with respect to some quantity other than time. In Fig. 4·34, the ball, which is constrained from tangential motion, is positioned radially on the surface of the disk by the motion r which is to be integrated. A rotation x of the disk rolls the ball, causing the cylinder to rotate in the

[1] Thomas M. Moore, "V-2 Range Control Technique," *Elec. Eng.*, C5, 303, July 1946.

[2] See, for example, M. Fry, "Designing Computing Mechanisms," reprinted from *Machine Design*, Penton, Cleveland, August 1945 to February 1946.

opposite direction. A given rotation of the disk causes a rotation of the cylinder proportional to this rotation times r, so that the total movement of the output y is a measure of the integral of r with respect to x.

The device operates correctly as r goes through zero and becomes negative. The output load should be small, or slipping will occur, with resulting error. But if it is carefully made and the load is negligible, the device can be very small (2 in. in diameter) and still give integration of less than 1 per cent error. Large integrators, such as those used in the MIT differential analyzer,[1] are accurate to 1 part in 10,000 and, with special care, 1 part in 100,000.

Electrolytic Integrators.—The amount of material removed from or deposited upon an electrode in an electrolyte is a very accurate measure of the time integral of the current through the electrode. Such a device is not practical as a continuous integrator because of the difficulty of obtaining a usable output proportional to the accumulation. The device is useful, however, where the desired output is not the integral, but a signal occurring when the integral has attained equality with some specified amount. In this case, the integration may start with a certain amount of plated material on one electrode; when all of this has left the electrode, a rise in voltage across the cell results, which can be utilized for the output signal.[2]

[1] V. Bush and S. Caldwell, "A New Type of Differential Analyzer," *Jour. Franklin Inst.*, **240**, 255–326, October 1945.

[2] Moore, *loc. cit.*

CHAPTER 5

THE GENERATION OF FUNCTIONS

BY D. MACRAE, JR. AND W. ROTH

5·1. Introduction.—In computer design it is often necessary to produce a prescribed function of an input variable. Such a function may have an explicit expression in closed form (for example, $\sin x$, x^3, $\sqrt{x^2 - x}$); it may be defined by an implicit relation; or it may be given simply as a set of points or a curve in two dimensions. If the function can be expressed in closed form, it is usually possible to devise a method of computing it by using the operations of arithmetic and calculus to express an equation that defines the function. Thus $y = \sin x$ might conceivably be produced[1] by instrumenting the equation

$$\frac{d^2y}{dx^2} + y = 0,$$

and $y = x^3$ by the equation

$$y = xxx.$$

It is also possible to produce some functions by the use of relations among physical variables that involve the functions directly. Examples of this are the projection of a rotating point on a straight line (resolvers, Scotch yoke) to produce sines and cosines, and the use of the low-current characteristic of a diode to produce exponentials or logarithms.

The use of simple defining operations or of relations among physical variables is in some cases difficult; and when the function to be computed is known only as a set of empirical values, these methods cannot be used. In such cases it is often convenient to use methods that are not peculiar to the function in question, such as the construction of nonlinear elements (potentiometers, gears, cams) or the combination of simple elements (for example, linear potentiometers) to approximate the desired functions. These will be called "curve-fitting" methods.

Some devices that produce nonlinear functions have only a single input and a single output; nonlinear gears and cams and certain vacuum-tube squaring devices are of this type. There are others that effectively have two inputs and may be used to multiply one of them by a nonlinear

[1] This method is used in the MIT differential analyzer; see Bush and Caldwell, "A New Type of Differential Analyzer," *Jour. Franklin Inst.*, **240**, 274, October 1945.

function of the other. Examples of these are resolvers and nonlinear potentiometers. These devices may be used as nonlinear modulating devices (Vol 19, Chap. 12, of the Radiation Laboratory Series) or, with the second input held constant, simply as generators of nonlinear functions.

Integration, differentiation, and combinations of these operations are also included as means of generating functions; for if a set of values of an input function is given, associated with values of an independent variable, these operations produce a corresponding set of output values.

CURVE FITTING

5·2. Construction of Nonlinear Elements.—The most straightforward way of producing a nonlinear function without solving an equation for it or using physical laws involving it is the construction of a nonlinear element. Elements of this type are potentiometers with nonuniform cards or unequally spaced wires, and gears and cams of varying radii; a function of two variables can be represented by a shaped surface. These elements can be used to approximate functions of any sort—analytic or nonanalytic—subject to limitations of construction, which may place bounds on the value of the function or its derivatives.

In the construction of nonlinear elements, the error is usually of a random nature and cannot be expressed conveniently as a function of the independent variable except by plotting an error curve for each particular element.[1] Production tolerances on these elements are usually expressed in terms of greatest permissible error (for example ± 0.1 or 1.0 per cent of maximum output or the corresponding number of thousandths of an inch). In some cases an estimate of probable error may be made.

Potentiometers.—In Vol. 17, Chap. 8[2] a number of general methods of winding potentiometers to produce desired resistance functions are given. Those which may be used for curve fitting are as follows:

1. Shaped mandrel. The length of a turn of wire is varied by winding on a mandrel or card of variable width, the width of the card varying linearly with the derivative of the desired function. If the card is smooth, there is a physical limitation on the maximum slope that it may have before turns roll off (about 15°). This places an upper limit on the second derivative of a function for which a potentiometer can be wound in this way. An approximation with straight-line segments is sometimes useful.[3]

[1] These correspond to the Class A errors defined in Sec. 2·6.

[2] See also Vol. 19, Sec. 12·23, of the Series for discussion of potentiometers.

[3] In Vol. 19, Sec. 12·26, an example of such a potentiometer is given. The application is a direct drive for a B-type radar display as a function of antenna position.

2. Controlled wire spacing. With wire of uniform resistivity, the spacing between wires can be varied to obtain a desired resistance function. One method of accomplishing this in quantity production is to use a servo winding device that equates the output resistance function to a standard during the winding process. The chief limitation of this method results from the minimum possible wire spacing; this puts an upper bound on the first derivative of a function that can be produced in this way. Figure 5·1 shows a function that has been wound[1] by this method. This is a hyperbolic function used in triangle solution (Sec. 6·4). On a card 6 in. in length, an accuracy of ± 0.25 per cent of resistance measured

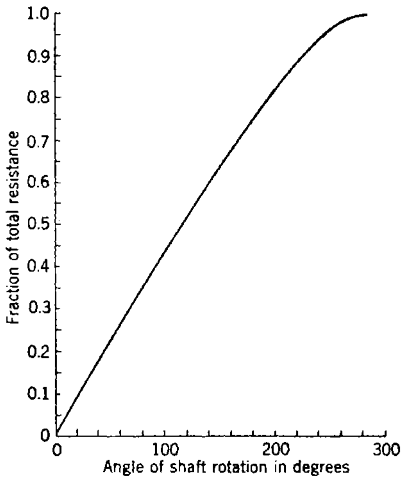


FIG. 5·1.—Function wound on a nonlinear potentiometer.

from the high-resistance end was attained. The total resistance was 20,000 ohms.

3. Control of wire size. An etching process may be used in connection with a servomechanism for winding nonlinear functions. Minimum wire size might be expected to be the chief limitation, again corresponding to an upper bound on the first derivative.
4. Step potentiometers. If discrete steps suffice in the production of the function, a device that is effectively a multicontact switch may be used, with resistors of appropriate values connected between the contacts.

[1] This work was done by the Thomas B. Gibbs Co., Delavan, Wis. Some of these potentiometers were also made by the Fairchild Camera and Instrument Corp., Jamaica, N.Y.

5. Tapped winding with shunts. If taps can be brought out from points on the winding, a resistance network can be connected to these points; in particular, shunting sections of the potentiometer with fixed resistors will produce a series of linear sections that may be used to approximate the desired function.

   The accuracy with which a nonlinear potentiometer can be wound depends on the nature of the function, the card length that can be used, and the method chosen for winding. For a function whose derivatives are not too great, it is possible to wind a potentiometer of 3-in. diameter with errors not exceeding 1 per cent of the maximum value of the function; with care, the errors of a 6-in. potentiometer can be held to 0.1 per cent. Some of the functions for which nonlinear potentiometers have been wound are the sine, cosine, secant, tangent, cotangent, and hyperbola.

6. Nonlinear mechanical elements. A mechanical motion varying nonlinearly with respect to an input shaft rotation can be obtained by the use of a cam.[1] A simple type of cam is a metal plate of variable radius that when rotated about an axis produces a variable radial displacement as output. A refinement of this type is the grooved cam, a plate with a spiral groove, which operates similarly but permits more than one full rotation of the input shaft. A function of two variables can be produced by the "barrel cam," which has a surface equivalent to a series of simple cams along the same shaft; distance along the shaft is then the second input variable.

   Another method of producing a shaft rotation that varies nonlinearly with an input shaft rotation is to use gears of variable radius.[2] There is a maximum value that the second derivative of a function may have if the function is to be represented in this way. For both cams and gears, a minimum value of the first derivative is determined by the radius of the shaft.

5·3. Nonlinear Functions with Simple Elements.—The difficulties of manufacture of special nonlinear elements make it desirable to use simpler and more readily available parts if possible. A large number of useful functions can be approximated by the use of a linear potentiometer with various resistance networks to produce a voltage that varies as a nonlinear function of potentiometer shaft rotation.

A greater variety of functions may be obtained if two or more potentiometer shafts are connected ("ganged") either directly or by gearing. In the design of networks of this sort, a combination of trial and error

[1] See "Basic Fire Control Mechanisms," Ordnance Pamphlet No. 1140, pp. 46*ff*.

[2] See Vol. 17 of the Radiation Laboratory Series.

with network analysis is necessary. The process of trial, or the selection of the general type of potentiometer network to be used, however, is facilitated by a knowledge of the sorts of functions that can be produced by simple potentiometer networks. For this purpose it is helpful to have families of output curves for various basic circuits. Once a circuit type has been chosen, the "best" values of the components may be found by an analytical method or by experimental measurements of the output function for different component values. In the latter case, a few experiments often show very rapidly the changes that can be produced in the function with variation of particular components. The question of whether analytical or experimental methods are to be used depends on the complexity of the network. If there is only one potentiometer in the network, an analytical method may be used to fit the desired function at three or four points; but if two or more potentiometers are used, an experimental method is sometimes preferable.

If the effect of a second potentiometer (or whatever element is used) is relatively independent of that of the first—the result being a sum, for example—the design process can be that of successive approximations. This method may be applied to either analytical or experimental design. An error curve is plotted for the approximation obtained with the first potentiometer, and another network is sought that will produce a curve nullifying this error. This resembles a convergent series expansion of the function. It can be applied to the Taylor expansion by using devices producing powers of the independent variable or to the Fourier expansion by ganging sine-cosine elements.

Loaded and ganged potentiometers may also be used to produce nonlinear resistance functions. Other examples of the combination of relatively simple devices to approximate desired functions are the design of shaped waveforms using exponentials,[1] and the use of linkages.[2]

A family of nonlinear functions may be generated by the use of a nonlinear element in a network with linear elements. A crystal rectifier, a lamp, or a thyrite element, for example, may be used with linear resistances. Another sort of characteristic that may be produced in this way is a resistance that is a nonlinear function of temperature; this has been done by combination of thermistors and resistors.[3]

When curve fitting is done by the combination of simple elements, a substantial portion of the error is usually predictable and independent

[1] Vol. 19, Chap. 8, of the Radiation Laboratory Series.

[2] Methods of approximating functions are treated extensively in Vol. 27 of the Series in connection with the design of linkages.

[3] R. Krock and N. Painter, "The Two-Disc D-c Thermistor Bridge Circuit," RL Report No. 502, Jan. 12, 1945. This report gives a fairly general theoretical treatment of curve fitting at three points by means of a nonlinear element.

of errors of the components used. This quantity is the difference between the desired function and the function that would actually be computed with perfect components.[1] It can be expressed in equations or graphs, and criteria can be stated for minimizing this error by proper choice of component values. A mathematical criterion such as the method of least squares may be used, but it is often too troublesome to use so refined a method. In general, a "weighting function" must be used if a given error in the nonlinear function produces different errors in the computer output, the errors depending on the region of the nonlinear function being used. If the weighting function is carried to an extreme, a set of points of the desired function can be chosen and the error minimized at those points only. If the number of points chosen is equal to the number of independently variable parameters of the approximating system, and if the desired function can be approximated with realizable components, these parameters may be so chosen that the error is zero at these points. This method, though it does not necessarily give the "best" fit, makes it possible to solve for component values in terms of the values of the function at the selected points.

5·4. Curve Fitting with Linear Potentiometers. *Analysis of Loaded Potentiometer.*—A method that is sometimes useful for approximating nonlinear functions involves the use of linear potentiometers in resistance networks. The most general resistive network containing a single potentiometer is a network including a-c or d-c generators and resistances, with the condition that two resistances having a common node constitute the potentiometer. These resistances will vary as x and $(1 - x)$ respectively, where x represents the angular displacement of the arm from one end of the potentiometer, expressed as a fraction of its full range. The common node corresponds to the contact made by the movable arm. It can be shown by means of the nodal analysis that any voltage appearing in the network can be expressed as a function of x in the form[2]

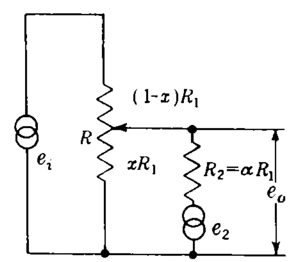


FIG. 5·2.—General loaded-potentiometer circuit with zero-impedance supply.

$$e_o = A\left(\frac{x^2 + a_1x + b_1}{x^2 + a_2x + b_2}\right), \tag{1}$$

where A, a_1, and b_1 are functions of the resistances and generator voltages and a_2 and b_2 are functions of the resistances only.

[1] This corresponds to the Class B errors mentioned in Sec. 2·6.

[2] A similar function is used for curve fitting with two rheostats; see N. Painter, "Matching Resistance Curves by Means of Two Linear Ganged Potentiometers and a Three Terminal Resistance Network," RL Report No. 610, Aug. 7, 1944.

A special case of some interest is that in which the potentiometer is supplied with voltage from a zero-impedance source and the output is taken at the potentiometer arm. This will be called a "loaded-potentiometer" circuit. In this case the network connected between the potentiometer arm and the reference node can be replaced, by means of Thévenin's theorem, by its internal resistance and open-circuit voltage. The circuit of Fig. 5·2 is the most general attainable subject to these restrictions. The circuit has three independently variable parameters that determine the shape of the output function: e_i, e_2, and $\alpha = R_2/R_1$. The level of the output may be changed by varying the potential of the reference node; this means that there are four independent variables instead of the five of the general case of Eq. (1). The function may also be moved along the x-axis by redefining the independent variable as a linear function of x. A rearrangement of the circuit for convenience of analysis is shown in Fig. 5·3. The two voltage sources e_i and e_2 with the corresponding series resistances $(1 - x)R_1$ and αR_1 can be replaced by current sources $e_i/(1 - x)R_1$ and $e_2/\alpha R_1$ respectively, with parallel resistances as shown in Fig. 5·4.[1] Since the current sources and resistances are all in parallel, the output e_o is simply the product of the total current and the effective parallel resistance, or

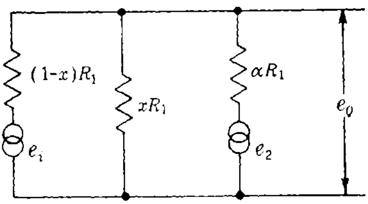
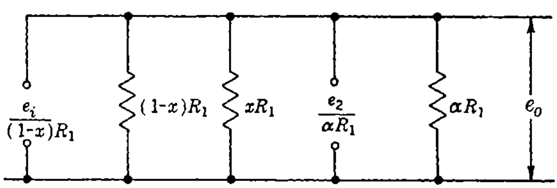


FIG. 5·3.—Rearrangement of loaded-potentiometer circuit.



$$e_0 = \frac{\dfrac{e_i}{(1-x)R} + \dfrac{e_2}{\alpha R_1}}{\dfrac{1}{(1-x)R} + \dfrac{1}{xR_1} + \dfrac{1}{\alpha R_1}}$$

FIG. 5·4.—Transformation of loaded-potentiometer circuit with current generators.

$$e_o = \frac{\dfrac{e_i}{1-x} + \dfrac{e_2}{\alpha}}{\dfrac{1}{1-x} + \dfrac{1}{\alpha} + \dfrac{1}{x}} = \frac{x[\alpha e_i + (1-x)e_2]}{\alpha + x(1-x)}. \tag{2}$$

[1] For this transformation see, for example, H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945, p. 12.

This can be generalized somewhat by the addition of fixed resistances at the ends of the potentiometer, in which case x varies over a smaller interval than $0 \leqq x \leqq 1$. This takes into consideration part of the effect of internal resistance of the voltage source; it does not mean, however, that the circuit of Fig. 5·3 is still the most general one when e_i has internal resistance; the Thévenin's theorem transformation by which that circuit

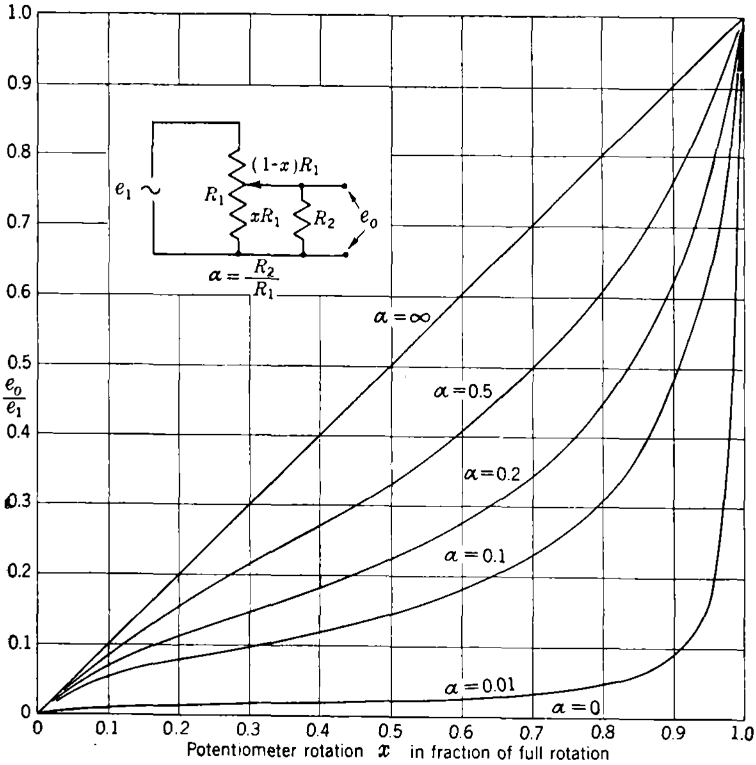


FIG. 5·5.—Output curves for loaded potentiometer with load connected to bottom end ($e_2 = 0$).

was obtained is not valid for this case, since the upper end of the potentiometer is no longer at a fixed voltage relative to the reference node. Some applications of Eq. (2) will now be considered.

Miscellaneous Functions from a Single Loaded Potentiometer.—In the process of fitting empirical functions with loaded potentiometers, it is helpful to know the sorts of function that can be produced with simple loading configurations. One such case is the circuit of Fig. 5·5, in which $e_2 = 0$. For this network Eq. (2) assumes the form

$$\frac{e_o}{e_i} = \frac{x\alpha}{\alpha + x(1 - x)}. \tag{3}$$

Figure 5·5 shows a family of curves for e_i/e_b as the resistance ratio α assumes different values. A corresponding family of curves may be

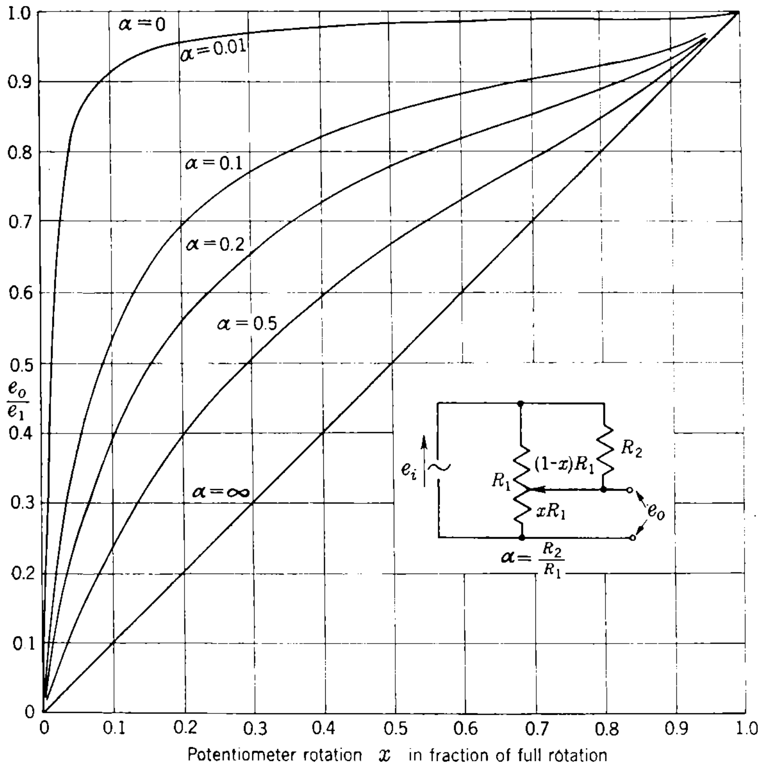


FIG. 5·6.—Output curves for loaded potentiometer with load connected to upper end ($e_2 = e_i$).

drawn for a load resistor connected to the upper end of the potentiometer. This corresponds to $e_2 = e_i$; the equation becomes

$$\frac{e_o}{e_i} = \frac{x(\alpha + 1 - x)}{\alpha + x(1 - x)}. \tag{4}$$

Curves of this function are shown in Fig. 5·6.

Figure 5·7 illustrates the effect of varying e_2 in the circuit of Fig. 5·2. If $e_2/e_i = \beta$, Eq. (2) becomes

$$\frac{e_o}{e_i} = \frac{x[\alpha + \beta(1 - x)]}{\alpha + x(1 - x)}. \tag{5}$$

Figure 5·7 shows a family of these curves for $\alpha = 0.1$ and various values of β.

Equation (5) may also be used to express the error term in the case when a potentiometer is being used to produce an essentially linear function but when a small amount of loading ($R_2 \gg R_1$ or $\alpha \gg 1$) is present.

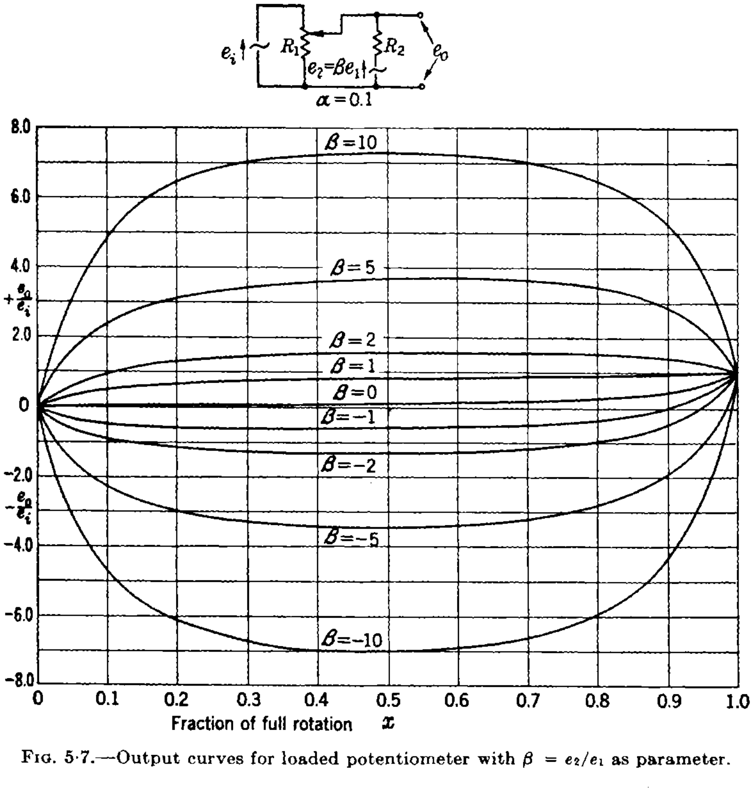


FIG. 5·7.—Output curves for loaded potentiometer with $\beta = e_2/e_1$ as parameter.

In this case Eq. (5) can be expressed in a more convenient form by dividing numerator and denominator by α.

$$\frac{e_o}{e_i} = \frac{x\left[1 + \dfrac{(1-x)}{\alpha}\beta\right]}{1 + \dfrac{x(1-x)}{\alpha}}.$$

This can be rewritten by means of the approximation

$$\frac{1+\epsilon_1}{1+\epsilon_2} \approx 1 + \epsilon_1 - \epsilon_2,$$

if $\epsilon_1, \epsilon_2 \ll 1$:

$$\frac{e_o}{e_i} \approx x\left[1 + \frac{(1-x)\beta}{\alpha} - \frac{x(1-x)}{\alpha}\right]. \tag{6}$$

The departure from linearity, expressed as a fraction of maximum output, is then

$$\frac{e_o - xe_i}{e_i} = \frac{x(1-x)(\beta - x)}{\alpha}.$$

This effect may be considered to be due to the varying output impedance of the potentiometer as a function of x. In the case $\beta = 0$, the magnitude of the error is $x^2(1-x)/\alpha$, which has a maximum of $0.15/\alpha$ at $x = \frac{2}{3}$. A plot of this function is shown in Fig. 5·8. One way of reducing the

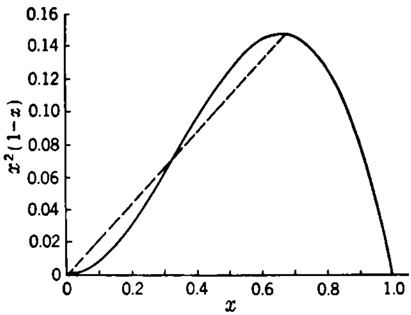


FIG. 5·8.—Potentiometer loading curve for $\alpha \gg 1$.

effective error is to operate over only a fraction of the full output voltage, as shown by the dotted line in Fig. 5·8; this may be done by inserting in series with the "high" end of the potentiometer a fixed resistance of about half the value of the potentiometer resistance and calibrating the system so that the output is correct at the high end of the potentiometer. The maximum output will then be only two-thirds the supply, but the maximum departure from linearity, expressed as a fraction of maximum output, will be reduced by a factor of about 6.

Another method of reducing the error somewhat is to make $\beta = \frac{1}{2}$ by connecting an additional load from the arm of the potentiometer to the high end. This effectively halves α. The maximum error in this case is $0.05/\alpha'$ (where α' = the new value of α) or $0.10/\alpha_0$ (α_0 = original value of α before addition of extra load). This method is inefficient in reducing the nonlinearity, relative to the first method.

For the approximation of symmetrical functions it is of interest to consider the form that Eq. (2) takes when the circuit is connected symmetrically with respect to the center of the potentiometer (as in Fig. 5·9). It will be convenient to let the independent variable be zero at the center of the potentiometer and the loading be symmetrical ($e_2 = e_1/2$) and to measure the output with respect to the level e_2. This can be done by defining a new input variable y such that

$$y = 2x - 1$$

and a new output variable u such that

$$u = e_o - e_2.$$

Substitution of these three conditions in Eq. (2) gives an expresson for u as a function of y,

$$u = \frac{4\alpha y e_2}{4\alpha + 1 - y^2}. \tag{7}$$

To put this into a more convenient form, let

$$4\alpha + 1 = \frac{1}{\delta}.$$

Then

$$u = e_2(1 - \delta)\frac{y}{1 - \delta y^2}. \tag{8}$$

The curves for this function resemble those of Fig. 5·5 but are symmetrical with respect to the origin.

Tangent Approximation.—By expanding the denominator of Eq. (8) in a power series,

$$u = e_2(1 - \delta)(y + \delta y^3 + \delta^2 y^5 + \delta^3 y^7 + \cdots). \tag{9}$$

The series for tan θ is

$$\tan \theta = \theta + \frac{\theta^3}{3} + \frac{2\theta^5}{15} + \frac{17\theta^7}{315} + \cdots .$$

By letting $\delta = \frac{1}{3}$ in Eq. (9) (that is, $\alpha = \frac{1}{2}$), the function produced by the loaded potentiometer becomes

$$u = \frac{2e_2}{3}\left(y + \frac{y^3}{3} + \frac{y^5}{9} + \frac{y^7}{27} + \cdots\right).$$

The quantity in parentheses differs from tan y by about $y^5/45$; at $y = \frac{1}{2}$ radian (29°) the error due to this term is only 0.04°. The circuit[1] is shown in Fig. 5·9. If $\alpha = 0.448$, the angular error can be held to $\pm 0.1°$ up to $y = 1$ radian.

To achieve in practice anything like the theoretical accuracy, it is necessary to calibrate the system carefully. This involves setting $e_2 = e_1/2$, zeroing the input shaft at the center of the potentiometer,

[1] Details of construction of such a circuit are given in G. D. Schott, "Loaded Potentiometer Triangle Solver," RL Group 63 Report, May 31, 1944. See also R. Hofstadter, "A Simple Potentiometer Circuit for Production of the Tangent Function," *Rev. Sci. Inst.*, **17**, 298–300, August 1946.

and setting $R_2 = R_1/2$. The condition $e_2 = e_1/2$ may be satisfied by supplying the potentiometer with alternating current by means of a center-tapped transformer. A phasing adjustment must be provided between the input dial and the potentiometer shaft in order that the zero position of the input shaft may correspond to the center of the potentiometer. The adjustment of R_2 may be made by means of a series potentiometer to produce the correct output at a known input.

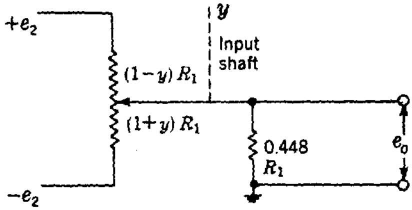


FIG. 5·9.—Circuit for approximating tangent function.

Secant Approximation.—A good approximation to the secant function can be obtained by the use of a potentiometer and auxiliary resistor (Fig. 5·10). In this case the output is not from the potentiometer tap, so the loaded-potentiometer analysis does not apply. By straightforward network analysis it is found that the ratio of output to input is

$$\frac{e_o}{e_1} = \frac{R_2}{R_2 + \frac{R_1}{4}(1 - y^2)} = \frac{4\alpha}{4\alpha + 1 - y^2}, \tag{10}$$

where y is the fractional angular displacement from the center of the potentiometer as defined in connection with symmetric functions and

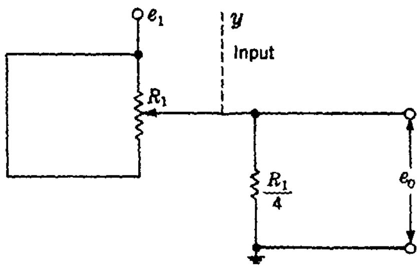


FIG. 5·10.—Circuit for approximating secant.

$\alpha = R_2/R_1$. This function is $1/y$ times the function of Eq. (7), for the same value of α. Again the substitution

$$4\alpha + 1 = \frac{1}{\delta}$$

facilitates the expansion of the denominator in a power series:

$$\frac{e_o}{e_1} = (1 - \delta)(1 + \delta y^2 + \delta^2 y^4 + \cdots).$$

The corresponding expression for the secant is

$$\sec y = 1 + \tfrac{1}{2}y^2 + \tfrac{5}{24}y^4 + \cdots$$

Let $\delta = \frac{1}{2}$($i.e.$, $\alpha = \frac{1}{4}$); the error then is approximately $\frac{1}{24}y^4$.

The same network, with the potentiometer as the bottom element and the resistor at the input, gives a fair approximation to a quadratic voltage function. The resistance of the potentiometer with ends connected is a quadratic function (Secs. 3·5, 5·9).

Square-root Approximation.—An approximation to the square root can be obtained with the circuit of Fig. 5·11. The output curve is compared with a square-root curve in this figure; the error is 1.7 per cent or less of maximum output over the range $0.04 \leqq x \leqq 1$.

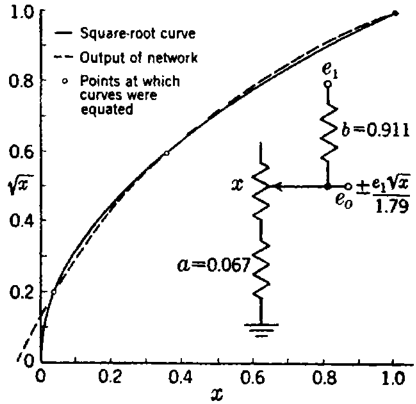


FIG. 5·11.—Approximation to square root.

The procedure in selecting the circuit constants is first to write the equation for the network type chosen,

$$y = \frac{e_o}{e_1} = \frac{A(x + a)}{x + a + b},$$

which contains the three unknown constants A, a, and b. This permits an exact fit to the curve $y = \sqrt{x}$ at three points (x_i,y_i). The points chosen were (0.04, 0.2), (0.36, 0.6), and (1.0, 1.0). Three equations in the three unknowns may be written, and a solution is possible by use of the relation

$$\frac{1}{y_i} = \frac{1}{A} + \frac{b}{A(x_i + a)},$$

the variable $1/A$ being eliminated by subtraction and b/A by division. The resulting equation, from which the dotted line in Fig. 5·11 was plotted, is

$$y = 1.79\left(\frac{x + 0.067}{x + 0.911}\right).$$

In the case of more complicated networks, for which the simultaneous equations are difficult to solve, fitting may be done more conveniently by plotting families of curves for the available and desired functions and superimposing them.

5·5. Other Combinations of Simple Elements. *Shaped Waveforms.*—The application of a simple waveform such as a positive step function to a network produces a response that may be shaped within certain limits by the choice of the elements in the network. If an RC-network is used, the resulting waveform will be a combination of exponentials. This method has been used to produce an approximation to a hyperbola.[1]

A waveform may also be approximated by a series of straight-line segments. This may be done by passing a linearly increasing current through a network of resistances and diodes;[2] transition from one segment to the next occurs when the waveform rises to a level at which a diode either begins to conduct or stops conducting, in either case changing the effective resistance of the circuit and the slope of the output waveform.

Linkages.—The use of mechanical linkages to approximate functions is a method having wide applicability (see Vol. 27). Bar linkages have been found to have certain advantages. Extremely long life may be expected from devices of this sort. Close mechanical tolerances must be held, however, the closeness depending on the accuracy desired. Linkages for functions of two variables have been designed by means of a method of successive approximations.

TRIGONOMETRIC FUNCTIONS

The trigonometric functions most often produced directly are the sine and cosine; other functions, such as the secant, tangent, and inverse trigonometric functions, may be produced by combining sine-cosine devices and by the use of feedback. The most common method of producing sines and cosines is the construction of a model in which the projection of a rotational motion on a line provides a quantity proportional to the cosine of the angle of rotation measured from the line. This principle is used in electrical resolvers, rectangular-card sine potentiometers, and Scotch yokes.

Trigonometric functions, like any other functions that vary sufficiently smoothly, may also be produced by curve-fitting methods; for example, nonlinear potentiometers have been used to produce sines and cosines, and cams to produce the secant. The choice in each case depends

[1] Vol. 19, Chap. 8, of the Radiation Laboratory Series.

[2] Ibid.

on whether the curve-fitting method or the use of physical laws involving sines and cosines is easier to instrument and more accurate. There are some devices that make use of both the physical law and a curve-fitting procedure. In some electrical resolvers having iron cores, for example, the model provides a first approximation to the sine function, but higher accuracy is obtained by careful spacing of stator windings or by the addition of a compensating cam, each of the latter operations being a case of curve fitting within a limited region.

5·6. Variation of Electrical Coupling by Rotation. *Types of Devices.* Both electromagnetic and electrostatic coupling coefficients can be varied mechanically. A typical electromagnetic device for producing sines and cosines consists of a cylindrical stator with one or more windings, inside which is a rotor with one or more windings.[1] Either air cores or iron cores may be used. The number of windings on the rotor and stator is referred to by calling the device a "1-to-3 phase" or "2-to-2 phase"

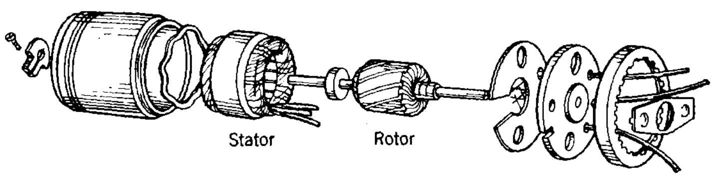


FIG. 5·12.—Bendix Autosyn, exploded view.

device, even though in ordinary use the a-c voltages on all these windings are essentially in the same time phase. Devices having 3-phase stators or rotors are usually called synchros; those with 2-phase stators or rotors are often called resolvers because they may be used to resolve a two-dimensional vector into its rectangular components. An air-core type of 2-to-2 phase device used for angle measurement is known as a goniometer.

Figure 5·12 shows an exploded view of a typical 2-to-2 phase, light-weight resolver. The drawing of Fig. 5·12 shows the cylindrical rotor, with slip rings for electrical information, and the stator into which the rotor fits. If the input is to the stators, there is no coupling between the stator windings unless current flows in the rotor windings.

Electrostatic coupling coefficients can be varied either by moving one condenser plate with respect to another or by moving dielectric material between the plates. A phase-shifting condenser that varies the phase of a sine wave by producing sines and cosines uses the latter method (see Vol. 17, Sec. 9·1, of the Series). This method is also used for transmission of angle data in the MIT differential analyzer.[2]

[1] The Radiation Laboratory Series, Vol. 17, Chap. 10, "Rotary Inductors."

[2] Bush and Caldwell, "A New Type of Differential Analyzer," *Jour. Franklin Inst.*, **240**, 4, 278, October 1945.

Uses.—An important application of electromagnetic devices of this sort is to the transmission of rotation. Three-phase devices (synchros) are used chiefly for this purpose. In this application little attention is paid to the production of sines and cosines as such; accuracy is measured in terms of angular errors rather than deviation of voltage from a sine function of rotation.

The same devices, but more commonly the 2-phase ones, may be used to produce voltages proportional to the sine and cosine of the angular position of the rotor. They may enter into a-c computing systems, in which case a sine-wave carrier is modulated by the rotor position; they may be used to produce sine and cosine components of triangular or trapezoidal waveforms for PPI sweeps (Vol. 22); they may be used with direct current and a specially built rotor to detect saturation resulting from d-c flux in null devices producing an angle output.

A 2-to-2 phase resolver may be used to rotate rectangular coordinates by resolving each of two inputs into components along directions defined by the two output coils; it may also be used, together with a servomechanism, to transform rectangular to polar coordinates (Sec. 6·3).

Either resolvers or phase-shifting condensers may be used to produce a phase shift varying linearly with shaft rotation. This results from combining sines and cosines. The two inputs are $\sin \omega t$ and $\cos \omega t$; these are multiplied within the device by $\cos \phi$ and $\sin \phi$ respectively and added to produce

$$\cos \phi \sin \omega t + \sin \phi \cos \omega t = \sin (\omega t + \phi).$$

Phase shifts of this sort are used in range measurement (Vol 20, Chaps. 5 and 6).

Theory of Operation of A-c Resolvers: Sources of Error.—The operation of an a-c resolver of the type shown in Fig. 5·12 may be analyzed similarly to that of an audio transformer. Differences lie in the variation of coupling coefficient with rotor position and the somewhat lower maximum value of coupling for the a-c resolver than for an audio transformer (for the Bendix resolver *XD*-759542, $k \approx 0.9$). An analysis in Vol. 17, Chap. 10, shows that, subject to certain assumptions, the input impedance is independent of rotor position. The frequency response of resolvers depends on the purpose for which they are designed; that of the Bendix resolver is shown in Fig. 5·13. Devices for resolving complex waveforms with high-frequency components may have a more uniform response or wider bandwidth.

Some of the errors of audio transformers as computing elements are found in the same form in resolvers. Nonlinearity of the iron core produces distortion of the output waveform and variation of stator-to-rotor voltage ratio with input voltage. For the Arma resolver (No. 213044) this effect causes errors in output voltage of the order of 0.14 per cent of maximum output for voltages from 0 to 80 volts rms at 400 cps. The effect of interwinding capacitance may also be observed if a resolver is operated with the output winding floating. The voltage ratio also varies with temperature, this variation being a function of the load as well.

One of the most important criteria of performance in synchros and resolvers is the deviation from sinusoidal output as a function of rotation. If the input is to a stator winding, it might be expected that the rotor would pick up a voltage measuring the projection perpendicular to the rotor winding of the magnetic field produced by the stator. In order for this to be true, the windings must be arranged to produce a suitable

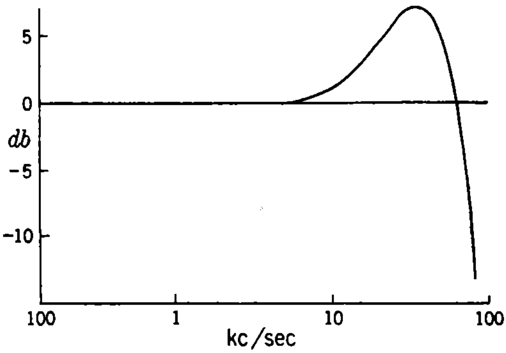


FIG. 5·13.—Approximate frequency response of Bendix resolver (stator-to-rotor ratio).

field. For any given arrangement of rotor windings, there are only certain field configurations for which a sine output can be obtained. Moreover, it is desirable that the field configuration be unchanged by the rotation of the rotor. This condition is fairly easily satisfied for air-core devices; but when there are iron cores, more care in design is necessary. Changes in the field may occur if the rotor is rotationally asymmetrical ("dumbbell" or "umbrella" type) or if at certain orientations winding slots of the rotor come opposite slots in the stator. This latter effect is reduced by designing the rotor or stator so that the slots of one are skewed with respect to those of the other (see Fig. 5·12).

If a load is connected to the output terminal of a resolver, certain errors may result. The input impedance of a stator will vary as a function of rotor position. Even if the input is supplied from a sufficiently low-impedance source, the output impedance will vary with rotation and the resulting variable loading will cause deviations from a sine function. For an Arma resolver, however, a 20,000-ohm load has negligible effect (<0.05 per cent) on the sine-function output; and for some purposes these resolvers have been connected in series[1] so that the output of one is the input for the next.

When two windings at 90° to each other are to be used either in producing the sine and cosine simultaneously or in producing the inverse tangent by a null method, errors may arise from the relation of one of these windings to the other. Their relative angle may differ from 90°; the maximum gain ratios from the two windings to the same pickup winding may differ; or the voltages on the windings may be slightly out of phase. The result of inequality of gains is an error that varies as the sine of twice the rotor angle; this may be remedied by the use of compensating resistors[2] or by the insertion of a gain control in the channel corresponding to one of the coordinates. If the two input voltages differ in phase, an error in output amplitude results. Effectively one of the input magnitudes is multiplied by the cosine of the phase difference. The

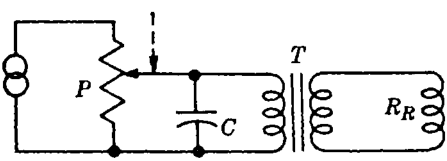


FIG. 5·14.—Transformer drive with amplitude control.

maximum fractional error in output amplitude due to a phase difference of ϕ radians between the input voltage is $\phi^2/2$ (for small ϕ).

Driving Circuits for A-c Resolvers.—Resolvers and synchros may be excited from an a-c line or from a transformer connected to an a-c line. This method is commonly used with synchros when the object is to transmit angular information. It may be used to produce sines and cosines but will not allow the general operation of multiplication of an input variable by sine and cosine.

A Variac[3] may be used to supply voltage to an input winding of a resolver; in this case the shaft rotation of the Variac provides the variable input to be multiplied by the sine or cosine. This method has the disadvantage of varying input impedance if it is desired to supply the Variac from any other source than a low-impedance line. Furthermore, it is difficult to attain high precision (0.1 per cent) with Variacs.[4]

If it is desired to excite an input winding of a resolver with voltage from a potentiometer having a resistance of several thousand ohms, an

[1] P. Weisz and B. Miller, "Transformation of Rectangular Coordinates Using Arma Resolvers," NDRC14-293, July 10, 1944.

[2] *Ibid.*

[3] See Sec. 3·11.

[4] H. S. Sack at Cornell University has done considerable work on precision Variacs (private communications). See also Chap. 3.

impedance-transforming device is usually necessary to prevent excessive loading of the potentiometer. Low output impedance is desirable in order that variation in input impedance with rotor position will not affect the voltage on the input winding. Such a device may be a step-down transformer, a cathode follower, or a more elaborate type of feedback amplifier.

A potentiometer may be used to vary amplitude of the voltage impressed on the rotor (Fig. 5·14). The potentiometer P is the amplitude control. In order to reduce the loading of this potentiometer, a stepdown transformer T is included. The input impedance at the primary of the transformer is essentially that of the reflected resolver-rotor impedance and thus can be quite large if a suitable value of step-down ratio is chosen. It should be noticed, however, that the potentiometer represents a variable internal generator impedance and together with the reflected inductance constitutes an RL series circuit in which the R is not a constant. This results in a variable phasing of the voltage impressed on the rotor, which cannot be tolerated in many applications. In order that a constant-phase voltage be impressed, the tuning capacitor C is included. This results in a resistive circuit in which the phase shift is independent of potentiometer setting. Since the Q of the circuit is low, the value of the tuning capacitor required is not critical. The usual commercial 10 per cent tolerance capacitors are suitable in most cases. This will, of course, restrict the frequency band over which the resolver can be used.

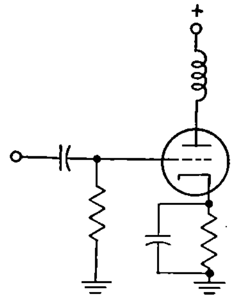


FIG. 5·15.—Plate-circuit driver.

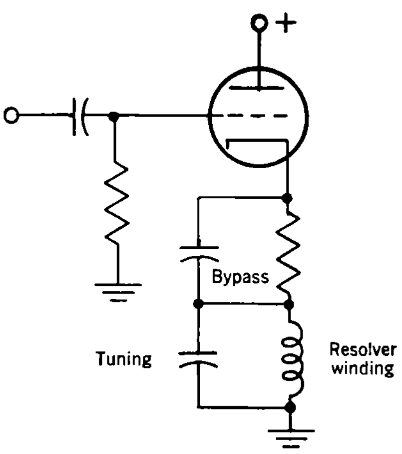


FIG. 5·16.—Cathode-circuit driver.

A simple type of vacuum-tube driver is an amplifier with the driven winding in the plate circuit (Fig. 5·15). This circuit provides amplification, but the gain varies with tube characteristic. If greater stability is desired, a cathode follower may be used, with the driven winding in the cathode circuit. A parallel tuning capacitor may be added to increase the range of operation. Cathode bias may be obtained by using a parallel RC in series with the load (Fig. 5·16). Both these circuits have the disadvantage that the d-c plate current passes through the driven winding, tending to saturate the core of the resolver.

Figure 5·17 shows two precision drivers for resolvers, discussed in detail in Vol. 18 of the Series. The circuit of Fig. 5·17*a* is essentially a cathode follower, but the d-c plate current is not allowed to flow through the winding being driven. With an Arma resolver stator as load, this circuit operates to 60-volt rms output with a maximum departure from linearity of 0.08 per cent; substitution of tubes causes variations in gain of ± 0.25 per cent. Figure 5·17*b* shows a two-stage driver circuit in which tube-substitution effects cause errors of 0.1 per cent or less, up to 20-volt rms output.

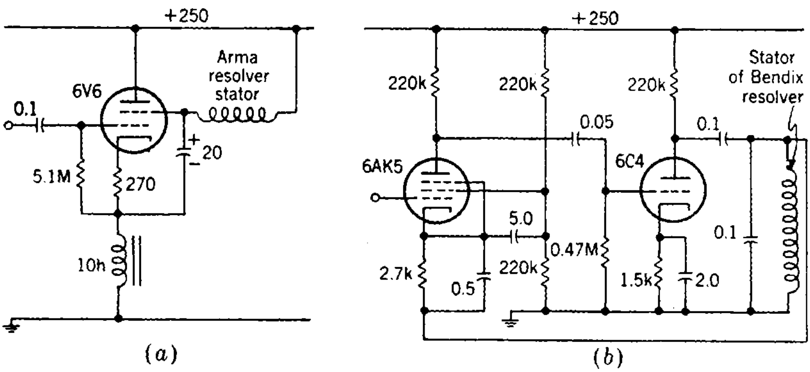


FIG. 5·17.—Precision driver circuits for resolvers. (*a*) Single-stage driver for Arma resolver; (*b*) two-stage driver for Bendix resolver.

If high accuracy is required, the output windings should operate into high-impedance loads, preferably grid circuits. Current flow due to low-impedance loading sets up corresponding fields within the resolver. These fields lead to interaction between the input windings in the case of 2-phase inputs (see Vol. 22 for information on sweep drivers).

D-c Resolvers.—A resolver[1] may be used to compute the inverse tangent of a ratio of two inputs (Sec. 5·12). If voltages measuring x and y are impressed on two stator windings, the resultant magnetic field is at an angle $\tan^{-1} y/x$ from the x-axis; the rotor may then be oriented perpendicular to the magnetic field by means of a servomechanism. This can be done fairly easily with alternating voltages (Sec. 6·9), for the error voltage picked up by a rotor winding may be used to actuate the servomechanism.

A similar operation is possible with d-c voltages representing x and y. It is necessary in this case to use a pickup element on the rotor that

[1] H. S. Sack *et al.*, Preliminary Report No. DCR-1, Cornell University, June 28, 1945.

indicates orientation in a steady magnetic field. Such an element is a magnetic core (Mo-permalloy or Mumetal, for example) excited by an a-c carrier and having two output coils connected in series opposition.[1] The output will be a second-harmonic signal which reverses in phase when the sense of d-c saturation of the core reverses. This signal may be fed to a phase detector that produces a d-c error signal for the servoamplifier.

The rotor of an Arma resolver was replaced with a d-c pickup element, and accuracy tests were made. For some orientation of the pickup element the effect of the earth's magnetic field was found to cause significant errors. If this effect was reduced either by proper orientation or by magnetic shielding, operation over the entire 360° was possible with errors not exceeding 0.5°.

5·7. Sine and Cosine Potentiometers.—Frequently it is convenient to employ a potentiometer to produce an output voltage proportional to the product of an input voltage and the sine of the angle through which the potentiometer shaft is rotated. If both the sine and the cosine functions are produced by the same potentiometer, the entire resolving operation has been performed; in some cases, however, it is necessary to use two independent potentiometers with the shafts displaced by 90° in order to develop both the sine and cosine functions. A discussion of sine potentiometers will be found in Vol. 17, Chap. 8, of the Radiation Laboratory Series.

Curve-fitting methods (Sec. 5·2) can be used to produce the sine function. One example is the construction of nonlinear potentiometers (using a shaped card, nonlinear wire spacing, etc.) which will produce a voltage or a resistance varying as the sine of shaft rotation. Such potentiometers are restricted in range to $\pm 90°$ or less, since the derivative of resistance with respect to rotation cannot change sign. If computer requirements necessitate a smaller range of angle, the maximum angle can be further restricted and the accuracy thereby increased.

Rectangular-card Sine Potentiometer: Theory of Operation.—If a contact is moved in a circle on the surface of a potentiometer card, the variation of potential with rotation is sinusoidal and unlimited rotation is possible. This principle is used in the rectangular-card sine potentiometer developed at the Radiation Laboratory; the construction of a typical sine potentiometer is shown in Fig. 5·18, and a schematic diagram in Fig. 5·19.

As is shown in Fig. 5·18, the card is of rectangular form with closely spaced wire wound continuously and may be rotated by the input shaft. Four brushes fixed to the case of the potentiometer and arranged geo-

[1] See Chap. 3 for discussion of magnetic amplifier.

metrically as shown assure proper contact pressure and alignment. Both sine and cosine functions may be obtained.

If E is the voltage impressed across the two ends of the winding and the dimensional constants are as shown in Fig. 5·19, expressions for the

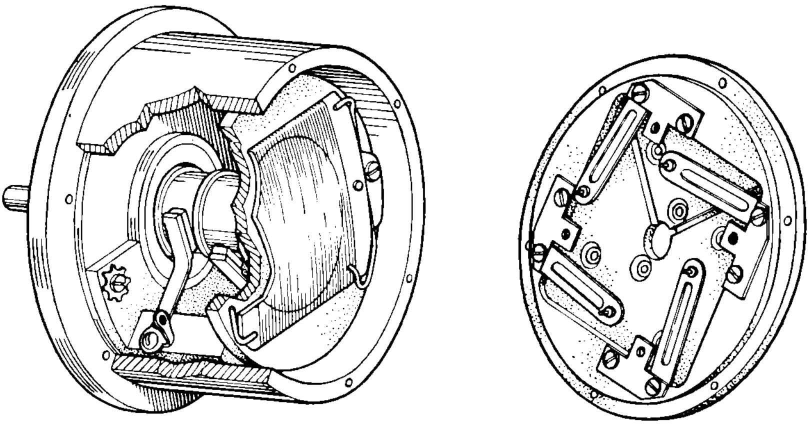


FIG. 5·18.—Sinusoidal potentiometer, type RL14. The performance is indicated by the following data: angular resolution $\pm \frac{1}{4}°$; life 5×10^6 revolutions at speeds up to 120 rpm; weight 1 lb 6 oz; diameter $4\frac{3}{8}$ in.; length $4\frac{1}{2}$ in.; winding resistance 32,000 ohms; applied voltage 300 volts or less; 305 turns per inch of 0.0025 =in. diameter nichrome wire; Formex insulation; brush force 1 oz; and resolution 0.3 per cent of peak voltage.

output voltages can easily be developed. It is to be noticed in particular that the axis of symmetry of the brushes does not necessarily correspond to the axis of rotation of the card. The significance of this will be discussed later.

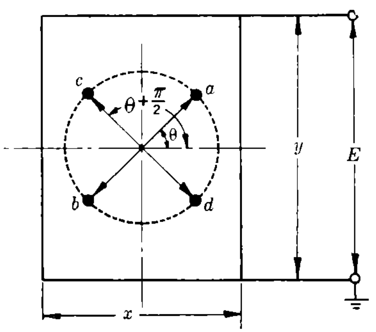


FIG. 5·19.—Schematic diagram of sine potentiometer. Points a, b, c, and d are brushes.

The card may be considered to have a potential gradient in the y-direction equal to E/y and in the x-direction of $E/2nx$ as a result of the potential drop along each turn (n being the total number of turns). The resultant of these, if the effect of wire resolution is neglected, is a uniform potential gradient at a slight angle to the y-axis. It is from this direction that brush rotation will be measured. Any brush moving in a circle relative to the card will then have a voltage with respect to ground that is the sum of a term varying as the cosine of the brush angle plus a constant term equal to the voltage at the center of the circle.

When the card is rotated, each brush in effect moves over the card in a circle whose center is the axis of rotation of the card. Thus if the voltage difference between any two brushes is taken as the output, the constant terms cancel and the difference is proportional to the cosine of the angle of rotation, measured from the angle at which the difference is a maximum. (For calibration purposes, greater accuracy can be obtained by referring angles to the position of zero output.) The two voltage differences between pairs of diametrically opposite brushes will be proportional to the sine and cosine of rotation if the four brushes are mounted accurately at the corners of a square. When the output differences are used in this way, the axis of rotation of the card need not correspond accurately to the geometrical center of the brushes.[1]

If the outputs from two of the brushes are used, without subtraction, as sine and cosine components, additional errors arise from lack of coincidence of the brush center and the axis of rotation. The angle subtended at the axis by the brushes may not be exactly 90°, so that the relation of the outputs is not that of sine and cosine; and the brushes may be at different radii from the axis.

The performance[2] of the RL14 sine potentiometer is indicated by the following data. When it is used as a resolver the angular accuracy is $\pm \frac{1}{2}°$ and the amplitude of the resultant vector is constant to within ± 0.65 per cent. If the voltages from single brushes are used, as mentioned above, the error may be several times greater. Even when the potentiometer is vibrated with an acceleration of 10g only slight evidence of noise is observable. The amplitude of this noise is less than $\frac{1}{2}$ per cent of the maximum signal. To remove this noise as well as brush noise, RC-filters are sometimes used at the output terminals.

The angle through which a brush must be rotated to go from one wire to the next varies from 0.17° when the brush is moving at right angles to the turns of wire to about 4° when the brush is moving parallel to the wire. The effective resolution of the potentiometer, however, is the angular resolution of the rotation of the vector whose components constitute the outputs, that is, the angle of the small but finite step with which the vector rotates. This angle varies from 0.17° when two of the brushes are moving normally to the wires to 0.24° when all four are at 45° with respect to the wires.

Input and Output Circuits.—Since the voltages from the two opposing brushes must be subtracted from each other in order to obtain the sine function, circuits associated with the potentiometer must be capable of

[1] A detailed discussion of the manufacturing problems involved in connection with such a potentiometer will be found in P. Rosenberg, "Sinusoidal Potentiometers Types RL10CB, RL10CD, RL10E, RL14," RL Report No. 423, Aug. 16, 1943.

[2] This material is taken from Vol. 19.

doing this. If high accuracy is to be obtained from a potentiometer having the desired linearity, it must be associated with output circuits having high impedance relative to the potentiometer resistance. In order to prevent loading, the usual practice is to connect the brushes directly to the grids of vacuum tubes. Figures 5·20 and 5·21 give two representative output circuits that are frequently used. If both sine and cosine outputs are desired, an output circuit is, of course, necessary for each component.

Figure 5·20 illustrates the type of circuit possible when a grounded voltage source is used. The voltage source can be either alternating or direct current in both Figs. 5·20 and 5·21. One end of the sine potentiometer is grounded as in Fig. 5·19. Thus the two output voltages must

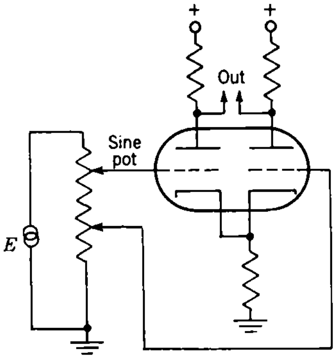


FIG. 5·20.—Sine potentiometer with grounded supply.

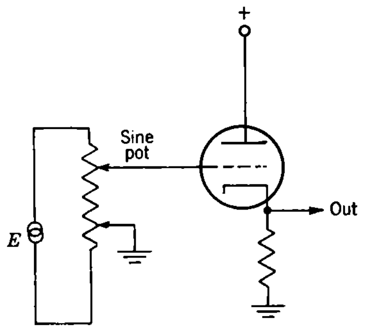


FIG. 5·21.—Sine potentiometer with floating supply.

be subtracted to eliminate the constant term. This can be accomplished by employing a so-called "difference amplifier" of which the circuit shown in Fig. 5·20 is an example.[1] The output leads are connected directly to the grids. The difference amplifier shown develops a balanced output; if this type of output is found undesirable, however, a difference amplifier developing an unbalanced or "single-ended" output can be used.

Figure 5·21 presents a typical circuit in which an unbalanced ("single-ended") voltage can be obtained directly from the potentiometer itself. As shown, this requires the use of a floating voltage supply, which, as above, can be either alternating or direct current. Such a supply might, for example, be a tachometer generator. Since the supply is not grounded, one of the two potentiometer brushes may be grounded, and the voltage measured on the ungrounded brush is the difference between

[1] Circuits for subtraction are discussed in Chap. 3 of this volume, and in Vols 18 and 19.

the two brush voltages. Only one component (either sine or cosine) can be obtained in this way. This single-ended output can be amplified in the usual manner if suitable precautions are taken against loading. The output circuit shown is a cathode follower. Since a high degree of linearity in the output stage is necessary in order for the over-all

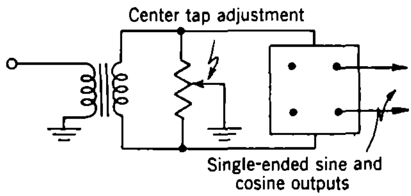


FIG. 5·22.—Push-pull a-c supply for sine potentiometer.

accuracy to be maintained, it is of advantage to employ output circuits with high gain and negative feedback.

Single-ended sine and cosine voltages can also be obtained by supplying the potentiometer in such a way that the center of the card is at ground potential. This may be done with push-pull a-c or d-c voltages

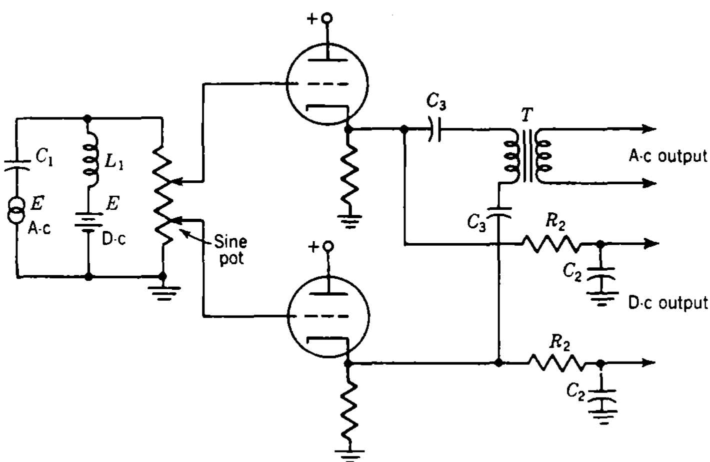


FIG. 5·23.—Method of using alternating and direct current simultaneously on sine potentiometer.

For alternating current a transformer may be used; if accurate adjustment of the center tap is desired, a control may be provided as in Fig. 5·22. For direct current an inverting feedback circuit may be used to produce a voltage that is the negative of a given voltage (Chap. 2).

It is sometimes necessary to obtain both a-c and d-c voltages that vary as the sine and cosine of an angle. The circuit diagram of Fig.

5·23 shows a method by which both signals can be obtained at the same time. The two separate voltage sources may be paralleled with isolation impedances in series with each to prevent alternating currents from flowing in the d-c supply and vice versa; or if such currents are not detrimental to the operation of the circuit, the supplies can be placed in series rather than parallel, without the isolation impedances.

The voltage from each brush is fed to the control grid of an isolation stage to prevent loading of the sine potentiometer by the filters. The a-c difference is obtained by connecting a high-pass filter between the outputs of the isolation stages; the d-c difference is obtained by connecting a low-pass filter between the same points. The filter designs are determined by the degree of separation desired and the phase shift or time lag permissible.

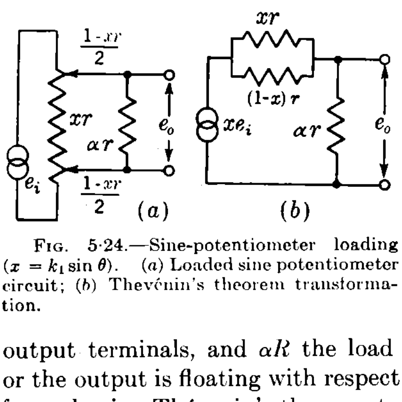


FIG. 5·24.—Sine-potentiometer loading ($x = k_1 \sin \theta$). (a) Loaded sine potentiometer circuit; (b) Thévenin's theorem transformation.

If a sine potentiometer is loaded by a resistance between the two terminals whose voltage difference is the output, the loading effect may be shown to be the same as given for a linear potentiometer in Eq. (3). Figure 5·24a shows this case. Let e_i be the generator voltage, R be the total resistance of the sinepot, xR the fraction included between the output terminals, and αR the load resistance. If either the generator or the output is floating with respect to ground, the circuit may be transformed using Thévenin's theorem to the form shown in Fig. 5·24b. The output e_o is given by the same equation as that for the linear potentiometer:

$$e_o = \frac{\alpha x e_i}{\alpha + x(1 - x)}.$$

For large α, the maximum error (expressed as a fraction of the voltage across the potentiometer) occurs at $x = \frac{2}{3}$ and is approximately equal to $0.15/\alpha$. A similar derivation may be carried out for the case of the single-ended output when the potentiometer is supplied from a center-tapped transformer of low output impedance; in this case the maximum loading error for large α is $0.10/\alpha$ times the maximum output.

5·8. Mechanical Methods.—Multiplication by sine or cosine of an angle is an operation that is done mechanically in many cases with suitable accuracy, although extreme accuracy (comparable to that obtainable with some electrical resolvers) requires very precise machining and considerable expense. Figure 5·25 shows an elementary form of crank and crosshead which can be used when a very crude sine or cosine function is permissible. The crank arm A is rotated by a drive shaft through an angle, and the output shaft B is actuated through the crankpin a by the connecting rod C; the displacement of the crosshead B is roughly proportional to the sine or cosine of this angle.

The departure of the output shaft displacement from a sine function is rather large unless the connecting rod is very long with respect to the crank arm.

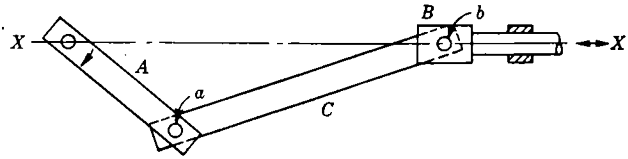


FIG. 5·25.—Crank and crosshead.

The Scotch yoke (see Vol. 17) is a similar mechanism with modifications that overcome the poor response of the simpler unit discussed above. Figure 5·26 illustrates a simple form of Scotch yoke. Again crank A is rotated through the given angle, and the output linear displacement along axis XX is proportional to the sine or cosine of this input angle. In this case, however, the response is mathematically correct if practical considerations such as tolerances, clearances, and surface smoothness are neglected. High accuracies are obtainable even with all these factors present, though at the expense of increased cost and difficulty of production.

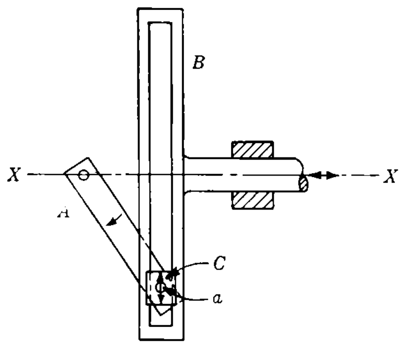


FIG. 5·26.—Scotch yoke.

The difference between the two units is that the Scotch yoke includes a slide and a channel that restricts the motion of the slide to an axis perpendicular to the axis of displacement XX. As the crank arm rotates, the slide C is displaced along the vertical axis, causing a horizontal movement of the output shaft B. The slide can be made integral with the XX channel unit in a compact mechanism. The inaccuracies in such units are those due to the need for clearance between the slide and the channel. This clearance introduces a backlash that in systems of high accuracies may be intolerable. Other less important points of error are the clearances necessary in the crankpin a and the bearings and so forth. In general, it may be said that the obtainable accuracy improves with the size of the unit, since with equal manufacturing methods the percentage clearance (actual clearance taken with respect to the size of the parts) decreases with increase in unit size.

Input angle information must, of course, be introduced mechanically so that the crank arm can be rotated through the desired angle. The displacement of the output shaft can be used in its mechanical form or may be converted into electrical information. An example of this change of representation is the use of the shaft to drive the arm of a linear potentiometer producing a voltage varying as the sine of input-shaft rotation.

Scotch yokes are commercially available. The more important factors that must be considered in evaluation of the relative merits of mechanical as opposed to electrical resolving schemes are accuracy obtainable with respect to accuracy desired, relative cost, size, weight, and associated electrical or mechanical units necessary for driving and/or obtaining output information. A general comparison of electrical and mechanical methods is given in Sec. 2·13.

5·9. Waveform Methods.—The sine wave, produced fairly easily by frequency-selective circuits, may be used to produce sine and cosine functions of variables other than time. Two methods may be mentioned, both of which make use of time selection (Vol. 19, Chap. 10, of this series). Probably the simplest in principle is to select the voltage corresponding to a given delay from the reference time and to store, or "remember," this voltage during the rest of the repetition interval. The input time delay may be generated from a voltage by means of a linear delay circuit; the entire operation, then, produces a voltage proportional to the sine of an input voltage.

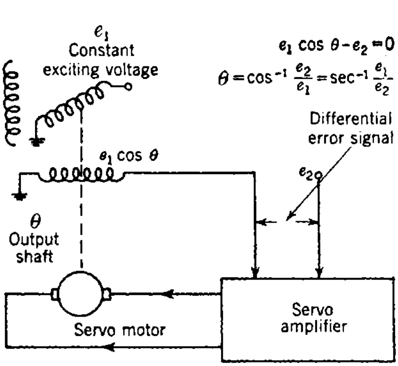


FIG. 5·27.—Device for producing inverse sine and cosine.

An alternative method has been developed[1] in which a variable-width rectangular gate is used to select a portion of a sine wave, and the average or integral is taken as the output. If the gate begins at a zero-voltage point of the sine wave, the integral measures the cosine of a quantity proportional to gate length.

5·10. Inverse Trigonometric Functions.—The inverse sine or cosine may be produced by connecting a sine-producing device and a servo

[1] B. Miller, P. Weisz, and W. F. G. Swann, "Circuits Using Resolvers and Coordinate Transformations by Means of Electrical Networks," NDRC Division 14 Report No. 288.

according to the methods of Sec. 2·5. An example of this operation in which a resolver is used is shown in Fig. 5·27. The voltage e_2, together with the resolver output $e_1 \cos \theta$, is fed into a differential servoamplifier that causes the servomechanism motor to turn the resolver rotor until the difference between the two amplifier inputs is zero. The shaft of the resolver rotor is then at an angle $\cos^{-1} (e_2/e_1)$ from its reference position. By redefining the reference position the inverse sine can be produced.

The inverse tangent may be produced by a similar servomechanism arrangement if both stators of the resolver are excited and the rotor is servoed to a null. Either an a-c or a d-c resolver (Sec. 5·5) can be used for this purpose. The circuit is shown in Fig. 5·28. The equation solved, in terms of the production of separate sine and cosine, can be written

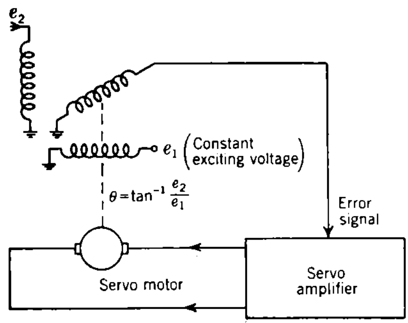


FIG. 5·28.—Device for producing inverse tangent and cotangent.

$$e_2 \cos \theta - e_1 \sin \theta = 0,$$

or

$$\theta = \tan^{-1} \frac{e_2}{e_1}.$$

This is discussed in more detail in connection with the transformation from rectangular to polar coordinates (Sec. 6·3). The inverse secant

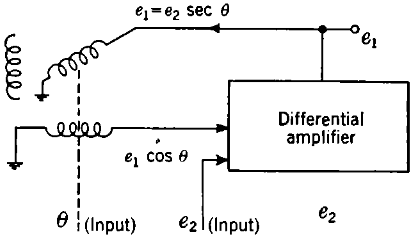


FIG. 5·29.—Device for producing secant and cosecant.

may also be produced with a resolver and servomechanism; the circuit of Fig. 5·27 will perform this operation provided only that e_1 becomes the variable voltage and e_2 the constant exciting voltage.

The secant itself is produced by feedback to the input winding rather than to the rotor shaft; Fig. 5·29 is a diagram of a device for producing the secant. A differential amplifier or equivalent device with high gain makes $e_1 \cos \theta = e_2$; as a result, $e_1 = e_2 \sec \theta$. By measuring θ from a different reference angle, the cosecant may be produced.

The method of amplitude comparison can be used to produce inverse sine or cosine delays from sine waves if the reference time corresponds to the proper fixed phase of the sine wave.

MISCELLANEOUS FUNCTIONS

5·11. Powers and Roots. *Powers.*—The operations of multiplication and integration may be used to produce powers of an independent variable (Vol. 19, Chap. 19 of this series). A straightforward way of producing a voltage proportional to the square of an input shaft rotation is to perform successive multiplications by means of two ganged potentiometers, the output of the first being fed through an isolating stage to the other (see Fig. 5·30).

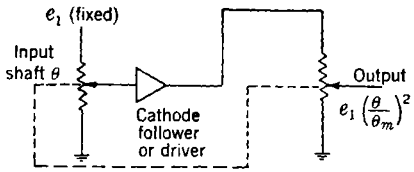


FIG. 5·30.—Ganged potentiometers for squaring. The maximum value of θ for the potentiometers is θ_m.

Integration may be used conveniently when the independent variable is time; for example, two successive integrations of a constant voltage may be used to produce a voltage that is a parabolic function of time (Vol. 19, Chap. 8, of this series). Squaring can also be done by averaging (integrating) triangular waveforms of variable duration or amplitude or by averaging rectangular waves whose amplitude and duration are varied together by the same input signal.

A precise mechanical device for squaring (capable of accuracy of 0.01 per cent) is the cone-cylinder combination produced by the Librascope Company (Vol. 17 of this series). In this device a wire is wound on a cone and a cylinder; and as the cone rotates, the rotation of the cylinder is determined by the length of wire transferred. This length is proportional to the square of the angular rotation of the cone, since the turns of wire are uniformly spaced along the axis of the cone, and the radius is therefore proportional to the angle through which the cone has rotated. The device may be considered an integrating type of squarer, for the length of wire l unwound from the cone or wound on the cylinder is

$$l = \int r \, d\theta,$$

where r is the radius of cone or cylinder and where θ is the angular rotation. For the cylinder, $r = r_0$, a constant; for the cone, $r = k\theta$. The fact that the same length of wire that is unwound from one is wound on the other is, in effect, an "equal sign," so that

$$\int r_0 \, d\theta_1 = \int k\theta_2 \, d\theta_2,$$

where θ_1 is the rotation of the cylinder and θ_2 is the rotation of the cone or

$$r_0\theta_1 = \frac{k\theta_2^2}{2}$$

if the initial conditions are properly chosen. Either θ_1 or θ_2 may be an input, so the device may be used either for squaring or for the extraction of square roots.

For low-precision applications (errors of 5 or 10 per cent) parabolic functions may be approximated satisfactorily by the characteristics of certain vacuum tubes (Vol. 19, Chap. 19, of this series).

Vacuum-tube characteristics may be used to greater advantage if a push-pull method is employed to cancel out odd-order terms in the Taylor expansion of the characteristic (Vol. 19, Chap. 19, of this series).

A quadratic function of rotation may be produced by short-circuiting the ends of a potentiometer together; the resistance between the arm and this common terminal is then

$$R = \frac{R_0}{4}(1 - y^2),$$

where R_0 is the potentiometer resistance and y is the motion of arm from center of potentiometer as a fraction of its peak positive or negative excursion, that is, limits of rotation correspond to $y = \pm 1$. This resistance element is used in approximating the secant function (Fig. 5·10); it is also used as a squaring element in triangle solution (Sec. 6·3).

Powers and roots, including nonintegral exponents, may be produced by means of logarithmic devices.

Roots.—Any multiplying or integrating device may be used to produce powers, and the techniques discussed for producing inverse functions may be used to produce roots. The cone-cylinder device may be used for the production of square roots as well as squares. A parabolic waveform may be used with a coincidence device to produce a square-root function. Ganged-potentiometer devices may be used with servomechanisms to produce roots. Figure 5·31 shows such a device for producing the cube root. If θ is the angle of shaft rotation of the three ganged-potentiometer arms and θ_m is the maximum rotation of a potentiometer, then the voltage output is $e_1(\theta/\theta_m)^3$. The servomechanism rotates the shaft until this quantity is equal to the input voltage e_2; the shaft rotation is then proportional to the cube root of e_2.

An approximation to the square-root function by means of a linear potentiometer is given in Sec. 5·4.

5·12. Exponentials and Logarithms.—Exponential functions arise frequently in physics and are usually connected with differential equations in which a function is linearly related to its first derivative. An example in electronics is the dependence of diode current on electrode voltages at low currents; an example of an exponential waveform is that resulting from the application of a step function to an *RC*-network (Vol. 19, Sec. 11·5). The most obvious use of exponentials in computation is the use of one physical system obeying an exponential law as a model

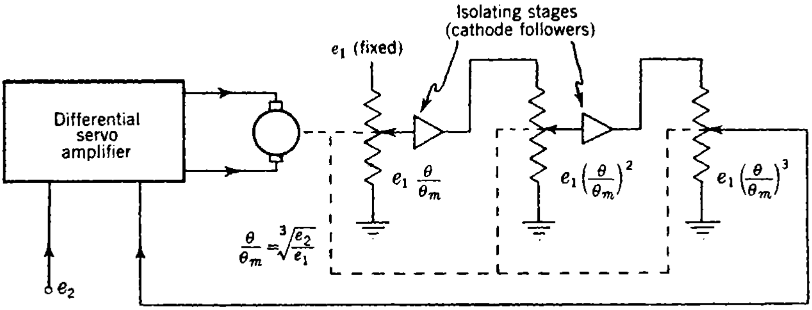


FIG. 5·31.—Device for producing cube root.

for another system, more difficult to construct, that obeys the same law. Thus if it is desired to compute automatically the attenuation of light intensity resulting from Lambert's exponential absorption law,[1] an exponential waveform generator might conceivably be used together with a linear delay circuit and a clamping circuit to change the representations of the input and output respectively.

If the physical system is such that the independent and dependent variables can be interchanged, the inverse of a function can be produced. This method has been used for the production of logarithms from exponential diode characteristics (Vol. 19, Chap. 9, of the Series) and might be used to produce logarithms from exponential waveforms. The automatic production of logarithms by electronic means is likely to have considerable application because logarithms can be added and put into an exponential producing device, the result being multiplication. The ability to perform a multiplication in the order of a few milliseconds is of

[1] $I = I_0e^{-\alpha x}$, where I = intensity at distance x;
I_0 = intensity at $x = 0$;
α = absorption coefficient.

considerable significance in computers involving combinations of operations, as will be pointed out in Chap. 6. At the present writing the chief multiplying device that has found wide application is the potentiometer, in which one of the inputs must be a shaft rotation. This limits the speed of computation considerably, since the time required to orient the shaft must be approximately 10^{-1} sec.

Contact Rectifiers.—It has been observed that the impedance of certain metal contact rectifiers[1] follows the law $E = r \log i$ over a substantial range of their forward characteristics. Commercial copper oxide rectifiers of various makes and sizes were found to follow that law consistently for a current range of 1/50 above a potential of about 0.070 volt per contact layer. Certain microwave mixer-type germanium contacts follow the logarithmic law through a current range of at least five decades, from the lowest measured value of 1 μa (with 0.16 volt) up to about 10 ma (0.50 volt) in any circuit and up to about 0.25 amp if the then appreciable current-proportional voltage drop within the semiconductor is canceled.

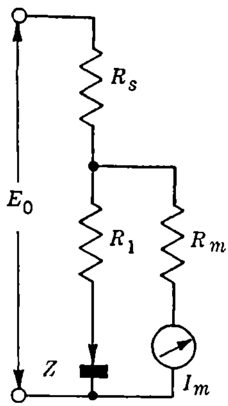


FIG. 5·32.—Voltage divider. R_1 = ohmic component of rectifier resistance; Z = logarithmic contact resistance; R = source resistance; R_m = meter resistance.

The simplest electrical circuit for obtaining logarithms is the voltage-divider circuit shown in Fig. 5·32. The output of this voltage divider is strictly logarithmic only (1) if the ohmic resistance R_1 within the rectifier is negligible compared with the logarithmic contact resistance Z, (2) if the source, or series, resistance R_s is large compared with Z, (3) if the load, or meter, resistance R_m is large compared with Z.

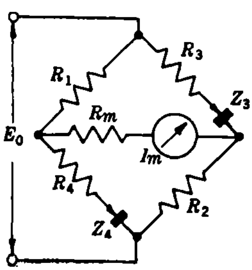


FIG. 5·33.—Bridge circuit.

An alternative circuit for the production of logarithms is the bridge circuit shown in Fig. 5·33. For convenience, symmetry will be assumed, with $R_1 = R_2$, $R_3 = R_4$, $Z_3 = Z_4$.

Nonlinear bridges of this type are known to have an output vs. input characteristic as plotted in Fig. 5·34 for negligible load current ($R_m \to \infty$). All curves were plotted with $R_1 = R_2 = 100$ ohms, $Z_3 = Z_4$ each two layers of 0.85 sq. in. copper oxide rectifiers in series, and with values of $R_3 = R_4$ varied from 0 to 100 ohms as indicated.

[1] This material is taken from H. E. Kallmann, "Three Applications of Nonlinear Resistors," RL Report No. Ja-5, Oct. 19, 1945, Part III. This report is to be published in *Electronics*.

The larger these resistances, which are in series with the rectifier contacts, the higher the input voltage at which the bridge balances ($E_m = 0$).

In particular, when $R_1 = R_2 = R_3 = R_4 = R$, the output voltage E_m no longer reverses. It then presents an exact replica of the voltage E_2 on the copper oxide rectifiers, rising logarithmically with rising bridge input. This result will be seen more clearly if the curve $R_3 = R_4 = 100$ is replotted on a semilogarithmic scale as in Fig. 5·35. The straight line thus obtained indicates a logarithmic relation over an input range of 34 db.

An evident merit of the bridge circuit is that the residual resistivity R_i of the contacts can be allowed for in the adjustment of the resistances R_3 and R_4. As in the voltage divider circuit, the output of the bridge circuit departs from the desired law at the low-current end of the range when Z grows comparable to R. The bridge circuit

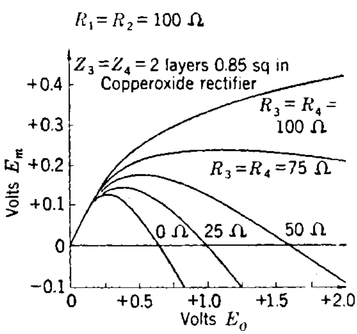


FIG. 5·34.—Characteristics of bridge circuit.

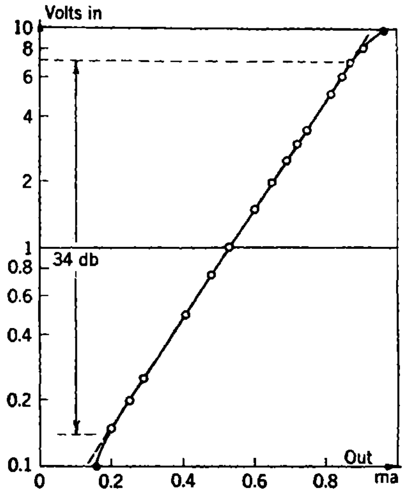


FIG. 5·35.—Logarithmic plot of bridge characteristic.

has the advantage that it may be operated at a much lower input voltage $E_{0_{\min}}$ without exceeding a specified error.

Logarithmic circuits lend themselves to combinations for the purpose of electronic multiplication or division. Fig. 5·36 shows, as an example, the combination of two log bridges. Their outputs are connected in series and fed, with opposite sign, to a meter, which thus reads the difference of their output currents. The current is then

$$i_{m_1} - i_{m_2} = A \log E_1 - A \log E_2 = A \log \frac{E_1}{E_2},$$

or proportional to the logarithm of the ratio of the two bridge inputs. The meter can thus be calibrated directly in decibels with a linear scale.

Two small resistors ρ with adjustable tap are shown in Fig. 5·36 for the initial adjustment of bridge balance. A model with each $R = 200$ ohms, each Z one layer of 0.85 sq in. copper oxide, $\rho = 5$ ohms, and a meter for 100 μa full scale with $R_m = 1800$ ohms was found reliable within ± 2 per cent of full scale for either input voltage varying from 0.14 to 7 volts, that is, up to ratios of ± 35 db.

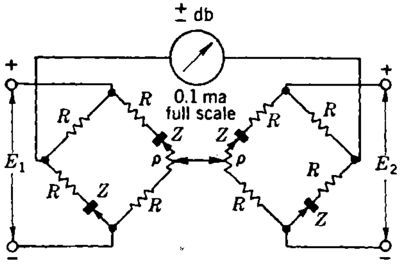


FIG. 5·36.—Combination of logarithmic bridges for division.

If circuits of this sort are to be used to obtain the product of rapidly varying waveforms, it is necessary to compute automatically the antilogarithm of the output current. This may conceivably be done by means of a third logarithmic bridge.

Diodes.—A diode limited to low-current values produces a variation of current with voltage that is very nearly an exponential. By the use

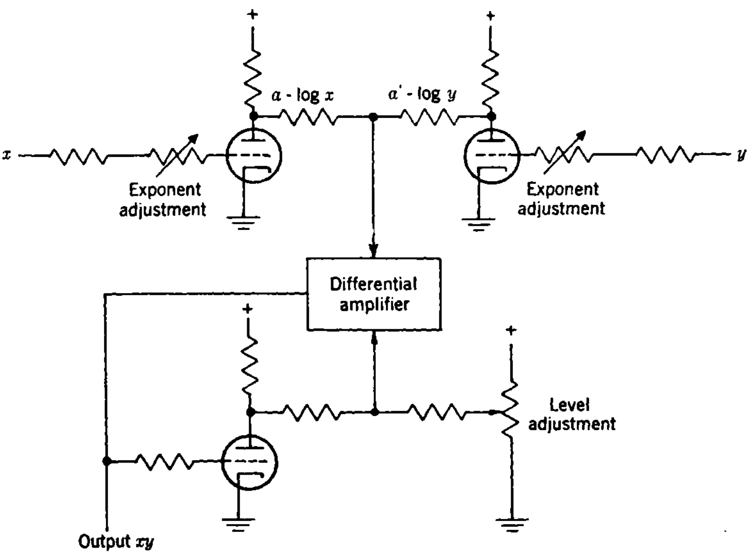


FIG. 5·37.—Use of logarithmic diode characteristics for multiplication.

of a large series resistor the current may be made the independent variable, the voltage output then being a logarithmic function of current. This property has been used in a multiplying device; the grid-cathode current characteristic of a triode is used to produce a logarithmic variation of grid potential, and the grid-plate characteristic to amplify this variation. Three such devices are combined in a feedback loop as shown in Fig. 5·37.

5·13. Integro-differential Functions.—As a generalization of the operations of differentiation and integration,[1] there are equations in which a combination of differentiation, integration, and addition is involved. Such, for example, are the equations of most passive *RC*-networks. The simple "differentiating" network (Fig. 5·38) has the differential equation

$$C \frac{d}{dt} (e_i - e_o) = \frac{e_o}{R},$$

which can be expressed in terms of the Laplace transforms as

$$\frac{\mathfrak{L}(e_o)}{\mathfrak{L}(e_i)} = \frac{pRC}{1 + pRC}.$$

A general expression for the ratio obtainable from a linear network is $f_1(p)/f_2(p)$, which can be reduced to the quotient of two polynomials.

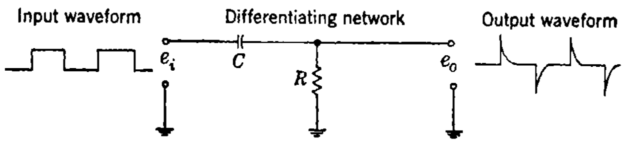


FIG. 5·38.—*RC* differentiating circuit acting on a rectangular waveform.

Uses.—In the problem of automatic range tracking (Chap. 8, Vol. 20) functions of this sort are used. A typical automatic range-tracking system consists essentially of two integrators and a feedback loop including the error-measuring device and producing the output range equal to that of the target. Considerations of loop stability, however, make it desirable to use a transfer ratio $(1 + pT)/p^2$ instead of the ratio $1/p^2$, which would correspond to two integrations. Additional factors $1/(1 + pT_s)$ are produced by the necessary data-smoothing networks, where T_s is the smoothing time constant. Similarly, in the design of stable high-performance servomechanisms or feedback amplifiers it may be desirable to use integro-differential functions of this sort in the feedback loop.

Another application of integro-differential functions is in the production of shaped waveforms. An easily available waveform may be fed into a function-producing circuit and a desired function or an approximation to it obtained.

Realizable Functions.—By means of passive *RLC*-networks a considerable variety of functions may be produced. In each case the function

[1] See Vol. 19, Chap. 2, and Vol. 18, Chap. 1, of the Radiation Laboratory Series.

can be expressed as a fraction whose denominator is a determinant and whose numerator is one of the minors of the determinant. Thus the numerator cannot contain higher powers of p than the denominator.

This restriction may be removed by the use of negative-feedback amplifiers. If a high-gain amplifier is connected as shown in Fig. 2·4 (Sec. 2·5) and if the admittances of the input and output networks can

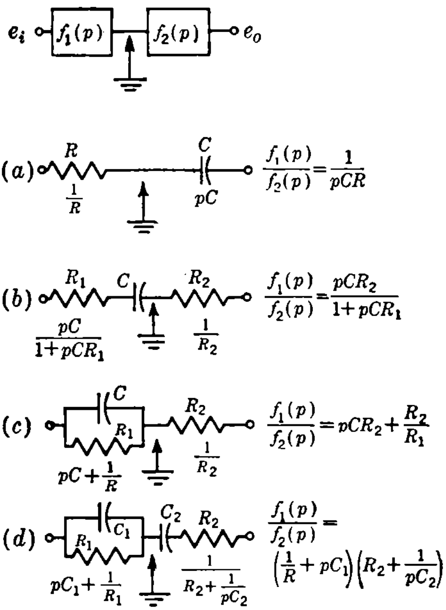


FIG. 5·39.—Transfer ratios of networks with feedback amplifiers.

be expressed operationally by $f_1(p)$ and $f_2(p)$ respectively, then the ratio of the transforms of output and input is very nearly

$$\frac{\mathcal{L}(e_0)}{\mathcal{L}(e_i)} = -\frac{f_1(p)}{f_2(p)}.$$

In this case, $f_2(p)$ may be made as simple as desired; if network (b) is a resistance, $f_2(p)$ is a constant. Several examples of circuits of this type are shown in Fig. 5·39. An arrow is used to indicate that the junction of the two networks is held at ground potential by the amplifier even though no current flows out at that point. Network (a) is that of a Miller feedback integrator. Network (b) produces a function that might as well be produced by a simple RC "differentiating" network; it is shown to illustrate the type of transfer ratio for which a passive network rather than a feedback amplifier should be used. Network (c) illustrates the principle that additive terms may be had by using parallel impedances in the input circuit; similarly, network (*d*) shows how additive terms are also introduced by series elements in the output circuit.

Realizable Output Waveforms.—If the input to a passive network is a step function, the output will be a combination of exponentials and damped sinusoids (Vol. 19, Chap. 11). This corresponds to the fact that the quotient $f_1(p)/f_2(p)$ can be expanded in a sum of partial fractions corresponding to the roots of the denominator. If a feedback amplifier is used with the network, the possibility of the root $p = 0$ enters; for a step-function input this corresponds to a linear output function, or, in the case of a double root, a quadratic function, etc. Thus the general function available in response to a step-function input is a linear combination of exponentials, sinusoids with exponential envelopes, and powers of t (time measured from the step). If more complicated functions are used as inputs, it is difficult to say in general what outputs are realizable, but the output can be calculated in any case from the input and the response to a step function by means of the superposition integral or convolution theorem.[1]

[1] Vol. 18, Chap. 1, of this series; or Gardner and Barnes, *Transients in Linear Systems*, Wiley, New York, 1942, pp. 228*ff*.

CHAPTER 6

GROUPED OPERATIONS

BY D. MACRAE, JR.

6·1. Introduction.—The computing devices to be considered in this chapter are intermediate in complexity between those of the preceding three chapters and the examples to be discussed in Chap. 7. The term "grouped operations" will include the production of a function or of a closely related pair of functions by combination of the separate devices described in the previous chapters.

Feedback techniques are important in the combination of operations. A particularly useful application is in the solution of simultaneous equations. An illustrative design of a single equation solver using feedback will be given.

Emphasis will be placed on applications of grouped operations to computation involving radar data. One such application is the automatic computation of the hypotenuse of a right triangle from the lengths of the legs; this is important in taking into account the effect of altitude in airborne radar equipment. Another problem, raised by radar navigation, is that of transforming one set of coordinates into another; several examples of this process will be given.

Problems Characteristic of Grouped Operations.—There are certain design problems that arise from connecting several devices. Combination of two computing elements may necessitate *change of representation* of the variables because the output of the first may not have the same representation as the desired input to the second. It is desirable to select devices for which the representations of outputs and inputs correspond, but other considerations such as availability and accuracy may make this impossible, so that a separate device for change of representation is often necessary. Two of the most common cases where change of representation is necessary are the multiplication of two voltages and the use of periodic-waveform computers with d-c or a-c voltage inputs. In the case of multiplication, a potentiometer may be chosen as the most convenient device available for the operation; in that case, a servomechanism must be used to convert one of the input voltages into rotation of the potentiometer shaft.[1] In computation using waveforms, the

[1] See, for example, Illustrative Design of an Equation Solver under Sec. 6·2.

input variables must often be converted to time delays by means of linear delay circuits.[1]

Consideration of *scale factors* is also necessary when operations are combined. Each separate device has limits of operation[2] for output and input; examples of such limits are the ends of potentiometers or the maximum output of an amplifier. In order to realize highest accuracy possible with given components, scale factors should be chosen so that the desired range of computation corresponds to these limits of operation. Thus, if the range of computation for a variable x, represented by voltage on a linear potentiometer, is to be $a \leqq x \leqq 2a$, the errors due to random departures of the potentiometer from linearity can be halved by inserting in series with it a fixed resistance of the same value as the potentiometer resistance. When two devices are connected, however, the two scale factors determined in this way may differ, and in order to use both to maximum advantage a scale-changing device should be inserted between them. For example, if a potentiometer having a rotation of 300° is coupled mechanically to a sine-cosine device whose shaft moves through only $\pm 30°$, the potentiometer error will be reduced if a 5/1 gear ratio is inserted between the two shafts.

Problems of *impedance* or *loading* are also raised by grouping operations. If a voltage representation is used, variation of input or output impedance may cause errors. Examples are the variation of output impedance of a potentiometer (Sec. 5·4) and the variation of input impedance of a resolver with a load on its output winding. These errors may be kept small if the output impedance of one device is much less than the input impedance of the following one. In order to satisfy this condition, it may be necessary to use an impedance-transforming device such as a cathode follower.

The *relative speed of computation* of devices that are to be used together is often of importance. The time required for the slowest device in the computer to produce sufficiently accurate results ("settling time") limits the rate at which the inputs may be changed; if the input varies sufficiently slowly to cause negligible change in output over a time equal to the settling time of the computer, these errors can be ignored. The problem of relative speeds is more serious, however, if a device in the computer has a minimum speed of computation. An electronic integrator (Sec. 4·7) is a device of this sort, for condenser leakage imposes a limitation on the maximum time for which the device may be operated accurately; if 1 per cent accuracy is desired, an operating time of the

[1] See, for example, Sec. 6·6.

[2] A general mathematical treatment of the interrelationship of limits of operation is given in Vol. 27 of the Series.

order of magnitude of an hour[1] may be realized. This limitation is important if it is desired to combine integration with multiplication. If a potentiometer is used as a multiplying device, it is necessary for computation to be relatively slow, for the mechanical information may have a settling time of the order of 0.1 sec; the integrator, on the other hand, must work relatively rapidly. These two requirements put fairly close limits on the time scale that can be used as regards rapid changes of the inputs. If instead of a potentiometer an electronic multiplying device with a settling time of 0.001 sec could be used, the integrator could be operated at a much greater speed, and the usable time scale would be much longer relative to the shortest interval over which accurate computation could be obtained.

Speed of response is also important in connection with the stability of computers using feedback loops. If a slow-operating element such as a servomechanism is in series with faster ones in a loop, the delay in the loop will be determined by the slow element. If slow and fast elements are in parallel feedback paths, the "loop gain" may be determined by one path for low frequencies and another for high frequencies.

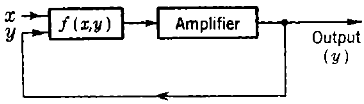


FIG. 6·1.—Block diagram for solution of $f(x,y) = 0$.

When several operations are combined, there are sometimes *alternate groupings* of the component devices from which the designer may choose. One of the most important choices is that of whether or not a given equation is to be solved by use of feedback.

6·2. Feedback and Implicit Functions. *Feedback vs. Direct Solution.*—A widely applicable method of solution using feedback (also discussed in Sec. 2·4) is shown in Fig. 6·1. The implicit function $y(x)$ is defined by the relation $f(x,y) = 0$. A method of producing $f(x,y)$ is devised by combination of operations of the sort described in Chaps. 3 to 5. This series of operations constitutes the block labeled $f(x,y)$. The output of this block, which is desired to be zero, is connected to the input of a high-gain amplifier. The output of this amplifier is fed back, with proper polarity to reduce the block output, as the y-input to the block $f(x,y)$. The result, if the design is satisfactory, is that the system will assume an equilibrium state in which the value of $f(x,y)$, the input to the amplifier, will be nearly zero. The higher the gain of the amplifier the more nearly will the equation $f(x,y) = 0$ be satisfied. The variable y is thus a solution of the equation to a high degree of approximation. There may be more than one independent variable or input; the variable

[1] This figure is subject to considerable uncertainty. The use of specially insulated condensers, special dielectrics, and compensation circuits (Vol. 19, Chap. 18) may permit longer operating times. See also Chap. 4 of this volume.

x in Fig. 6·1 may be considered to represent several such inputs. One possible application of this method is in finding real roots of polynomials with variable coefficients.

In order to determine whether or not a solution using feedback is desirable for a particular equation, the designer must compare the complexity of the feedback solution with that of a solution without feedback, where one is possible. In some simple cases a solution with feedback is obviously out of the question. For the equation $y = x_1 + x_2$, there is no point in using the implicit function $y - x_1 - x_2 = 0$ with a feedback loop of the type of Fig. 6·1. On the other hand, suppose that the

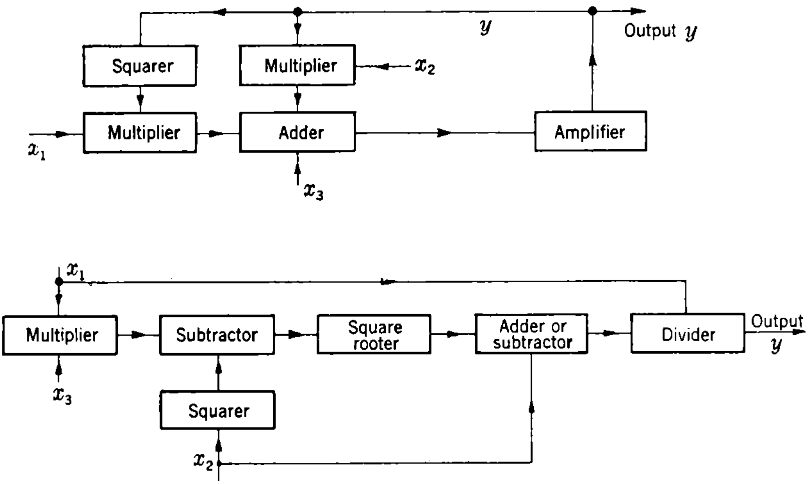


FIG. 6·2.—Block diagrams for solution of quadratic equation. (a) Solution of the implicit function $x_1y^2 + x_2y + x_3 = 0$ by feedback; (b) solution of $x_1y^2 + x_2y + x_3 = 0$ using the explicit function $y = \dfrac{-x_2 \pm \sqrt{x_2^2 - 4x_1x_3}}{2x_1}$.

equation to be solved is the quadratic $x_1y^2 + x_2y + x_3 = 0$, where x_1, x_2, and x_3 are inputs and y is the output. If the equation has a real root in the region of operation, a solution may be set up using a feedback loop of the type shown in Fig. 6·1. A block diagram of such a solution is shown in Fig. 6·2a. The output y is applied to a squaring device[1] followed by a multiplier, producing the function x_1y^2; it is also applied to another multiplier and an adder, producing $x_1y^2 + x_2y + x_3$. This is then fed back through an amplifier as described above.

In this case there is the alternative of a solution using an explicit function according to the binomial formula

[1] Squaring may be accomplished by any of the methods discussed in Chap. 5 or in Vol. 19, Chap. 19

$$y = \frac{-x_2 \pm \sqrt{x_2^2 - 4x_1x_3}}{2x_1}.$$

A block diagram of this solution is shown in Fig. 6·2*b*. The function is built up as follows: First x_2 is squared; the product $4x_1x_3$ is computed; and these quantities are subtracted (the constant factor 4 being taken into account by proper setting of scale factors in the multiplying and subtracting blocks); the square root is taken; the result added to or subtracted from $-x_2$; and this quantity divided by $2x_1$. The computation of the negative of x_2 is not included as a separate operation, because it has not been stated whether x_2 is positive or negative. Comparison of the two block diagrams shows that the solution with feedback requires fewer operations and that the corresponding device is probably easier to build. For cubic and quartic equations the complexity of the direct solution without feedback increases considerably; in the case of higher-

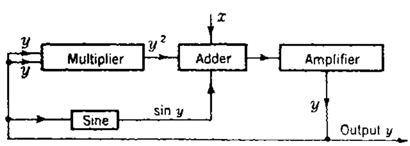


FIG. 6·3.—Block diagram for solution of $y^2 + \sin y + x = 0$.

order algebraic equations and many transcendental equations there is no explicit solution in terms of simple operations, so that if a solution without feedback is to be made, a special nonlinear device must be constructed by curve-fitting methods. This becomes difficult as the number of variables increases.

Illustrative Design of an Equation Solver.—To show how some of the design problems of grouped operations are met, an illustrative design process will be given for a computer solving the equation

$$y^2 + \sin y + x = 0.$$

Since this equation involves only two variables, the solution might be computed automatically by the construction of a suitable nonlinear potentiometer, cam, or similar element. The design to be discussed, however, does not use curve-fitting methods. A block diagram for the solution of this equation is shown in Fig. 6·3. It is constructed according to the method of Fig. 6·1; the functions y^2, $\sin y$, and x are added to produce the desired function of x and y; and this function is amplified and fed back. A corresponding schematic circuit diagram is shown in Fig. 6·4. The input x is represented by a shaft rotation, which is converted to an alternating voltage by a potentiometer. The amplifier output is an alternating voltage representing y; the representation of y is changed to a shaft rotation (which is taken as the output) by a servomotor and an amplifier that compares the y-voltage with a potentiometer output voltage. The y-shaft turns the rotor of a resolver, which is supplied with the same a-c voltage as is across the x- and y-potentiometers. The output of a stator winding of this resolver is proportional to sin y if scale factors have been properly adjusted. The y-voltage is multiplied by the y-shaft rotation by means of another potentiometer, producing y^2. The voltages representing the three functions x, y^2, and sin y are then added by means of the resistance network $R_1R_2R_3$, and the resulting voltage is the input to the amplifier.

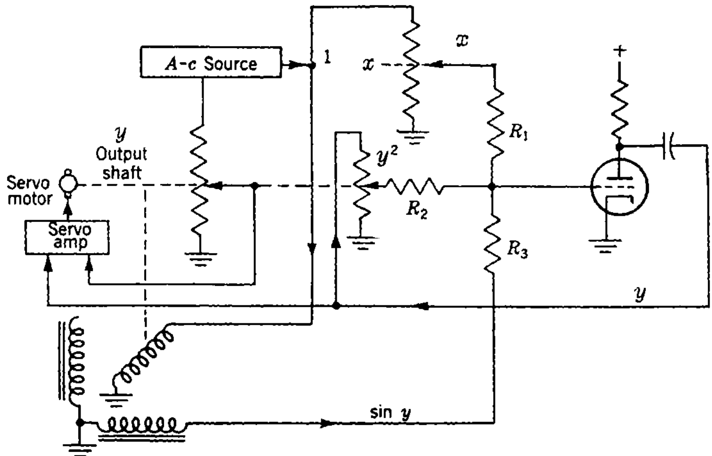


FIG. 6·4.—Circuit diagram of computer for $y^2 + \sin y + x = 0$.

Some preliminary design considerations may be mentioned in connection with this computation. One of the first factors to be considered is that of accuracy; without detailed calculation it may be said that the use of available resolvers (of 0.2 per cent peak output error) and potentiometers of 0.5 per cent peak deviation from linearity will permit computation with resulting probable errors not exceeding 1 per cent of the range of output.

Loading of the y^2- and x-potentiometers by the adding network must also be considered. If with available potentiometers and resistors this error is excessive, cathode followers may be necessary.

As an aid in determining scale factors and defining the necessary calibration adjustments, the equation solved by the system may be written including scale factors. Suppose that x and y are defined by dials on the respective shafts. The dials must be "phased" with respect to the shafts so that when the dials read zero, all potentiometers and the resolver are at positions of zero voltage output. This implies four phasing adjustments. If it is assumed that these have been made, the equations that determine the scale factors of the system can be written. Let the shaft rotation representing x be $\theta_x = k_1x$ and that representing y be $\theta_y = k_2y$. Assume that all three potentiometers have the same maximum rotation θ_m (if they do not, the inequalities can be compensated by adjustment of R_1, R_2, or R_3). Let E represent the a-c supply voltage. It will be convenient to treat the resistance network in terms of the conductances $G_1 = 1/R_1$, $G_2 = 1/R_2$, $G_3 = 1/R_3$,

$$G = G_1 + G_2 + G_3.$$

The servomechanism sets the amplified error signal equal to $E(\theta_y/\theta_m)$. The three voltages entering the resistance network (R_1, R_2, R_3) are $E(\theta_x/\theta_m)$, $E(\theta_y/\theta_m)^2$, and $E \sin \theta_y$. The resolver is assumed to have unity maximum voltage ratio; if not, this can be taken into account by adjusting R_3. The equation solved is then

$$\frac{G_1}{G} E \left(\frac{\theta_x}{\theta_m}\right) + \frac{G_2}{G} E \left(\frac{\theta_y}{\theta_m}\right)^2 + \frac{G_3}{G} E \sin \theta_y = E \frac{\theta_y}{\theta_m} \frac{1}{A},$$

where A is the gain of the amplifier (see Vol. 19, Chap. 18 for analysis of adding network). The factor E may be removed from the equation by dividing through; the solution is therefore independent of variations in E within the limits for which the above equation holds. Simplifying,

$$\frac{k_1x}{R_1} + \frac{(k_2y)^2}{\theta_mR_2} + \frac{\theta_m}{R_3} \sin k_2y = \frac{Gk_2y}{A} \approx 0.$$

The scale factor calibration then consists in setting

$$\frac{k_1}{R_1} = \frac{k_2^2}{\theta_mR_2} = \frac{\theta_m}{R_3} \qquad \text{and} \qquad k_2 = 1,$$

in order to make the equation assume the desired form. The first two conditions may be satisfied by adjustment of two of the resistors R_1, R_2, and R_3; this may be done conveniently by potentiometers in series with fixed resistors. These adjustments are made in such a way as to produce correct output values (y) for given values of the input (x); by this method all components contributing to the scale factors are taken into account. The condition $k_2 = 1$ can be satisfied (if y is measured in radians) by using a y-dial marked in radians or degrees. When calibrated the system solves the equation

$$\frac{k_1}{k_2GR_1} (y^2 + \sin y + x) = \frac{y}{A}.$$

The Solution of Simultaneous Equations.—In three-color printing it is of importance to be able to solve automatically a set of three simultaneous equations of the type

$$
\begin{aligned}
X = {} & (1 - c)(1 - m)(1 - y)X_w + c(1 - m)(1 - y)X_c \\
& + m(1 - c)(1 - y)X_m + y(1 - c)(1 - m)X_y + my(1 - c)X_{my} \\
& \qquad + cy(1 - m)X_{cy} + cm(1 - y)X_{cm} + cmy\, X_{cmy}, \\
Y = {} & (1 - c)(1 - m)(1 - y)Y_w + \cdots, \\
Z = {} & (1 - c)(1 - m)(1 - y)Z_w + \cdots,
\end{aligned}
$$

where X_w, X_c, Y_w, etc., are constant coefficients, X, Y, Z are known quantities, and c, m, and y are the unknowns.

The principal problems involved are the electrical representation of terms of the type

$$(1 - c)(1 - m)(1 - y)X_w \text{ or } cm(1 - y)Y_{cm},$$

and the construction of a feedback system whereby all three equations may be solved simultaneously. A multiplying device based on a probability principle was developed by Prof. A. C. Hardy of Massachusetts Institute of Technology for this purpose.[1] He then combined a number of these devices to produce the indicated functions that will be denoted by X', Y' and Z'. The differences, or error signals, $X - X'$, $Y - Y'$, and $Z - Z'$ were amplified and fed back as the three variables c, m, and y. The question of which variable is to be derived from which error signal can be determined by an inspection of the partial derivatives of X', Y', and Z' with respect to the three variables, and the choice can be made according to where the feedback would be most effective, that is, where the partial derivative is greatest. A device was constructed to solve these equations, and results accurate to within 4 per cent or better of maximum output were obtained. The entire solution could be done within approximately 500 μsec. The repetition frequencies for the rectangular waves used in multiplication were approximately 20 kc/sec.

A theoretical analysis of the stability of a device of this sort seems difficult. It may be said, however, that the stability conditions are probably no more severe than for three simultaneous linear equations. The reason for this is that the three third-order equations given above can be replaced by equivalent linear equations involving the partial derivatives of X', Y', and Z', with respect to the variables c, m, and y. These equivalent linear equations will describe adequately the behavior of a system within a small region in the neighborhood of a solution.

THE SOLUTION OF RIGHT TRIANGLES

6·3. The Problem of Right-triangle Solution.—In computers for use with airborne radar equipment, it is often necessary to find the hypote-

[1] Unpublished work, described at MIT Physics Colloquium, Jan. 31, 1946. See Vol. 19, Sec. 19·5, and Sec. 3·17 of this volume.

nuse of a right triangle when the lengths of the legs are given. This is because the range information provided by the radar is "slant range"; that is, distance along a straight line from the aircraft to the object giving the echo. It is more convenient in a computer, however, to use ground range (distance from a point on the ground beneath the aircraft to the object), for ground-range rates are much more nearly constant than slant-range rates. This is a very desirable feature if rates are to be found by tracking (Vol. 20, Chap. 7, of the Series). The procedure in computer design, therefore, is to do whatever rate computation is necessary in terms of ground range and then to find the hypotenuse of the altitude triangle as the last step before comparing computed range with observed radar range. In this section the term "triangle solution" will denote the process of finding the hypotenuse of a right triangle, but not of finding angles or of solving non-right triangles. The following symbols will be used:

$$
\begin{aligned}
r &= \text{ground range}, \\
h &= \text{altitude}, \\
s &= \text{slant range}.
\end{aligned}
$$

The problem of finding the hypotenuse of a right triangle is not, of course, confined to the altitude triangle in radar computers. Since the triangle solvers to be described were designed for radar problems, however, it will be seen that some of them have features which particularly fit them for use with radar. One such feature is the representation of quantities by the time delay of pulses from a periodic reference pulse; time-delay representation or time modulation (Vol. 19, Chap. 13 of the Series) finds more use in triangle solvers than in other types of computers because the triangle solver often directly precedes the time-modulation circuit. Another feature desirable in a triangle solver is that it incidentally provide an h-delay output, for the aircraft's altitude may be determined by comparing a variable-delay marker on the radar display with the first ground return on the radar, which is delayed from the transmitted pulse by an amount measuring h. If this method is used, some of the errors of the ground-range determination tend to be canceled out, particularly the zero error due to delays in the radar. It is tacitly assumed that the altitude above sea level of the object giving the echo is the same as that of the ground beneath the plane; if it is not, inherent errors result.

One class of triangle solvers which will be discussed may be called "algebraic" devices. These include those which solve the equation $r^2 + h^2 = s^2$ by squaring, adding, and extracting the square root. They also include a device that uses the relation $(s - r)(s + r) = h^2$. Another class of triangle solvers which will be mentioned briefly consists of those devices which make use of special nonlinear elements constructed by curve-fitting methods (Sec. 5·2). A third class of devices that may be used for triangle solution includes those which transform rectangular to polar coordinates (Sec. 6·9). Finding the hypotenuse of a right triangle is simply a special case of this transformation in which the angle output is not used.

Two triangle solvers will be discussed in detail. One, a parabolic-waveform computer developed in England, is of particular interest for its application of shaped waveforms to computation, as well as for its economical combination of component circuits. The other, a phase-shift triangle solver using a feedback integrator and differentiator, is economical in use of parts and contains interesting provisions for canceling out certain errors.

6·4. Algebraic Methods of Right-triangle Solution. *Block Diagram of Squaring Method.*—A general diagram of the squaring-type triangle solver is shown in Fig. 6·5. The quantities r and h are squared and added, and the result is applied to a device that extracts square roots. Depending on the sort of individual blocks available, the extraction of the square root may be done either directly or by the use of a squaring device with feedback. If squaring devices employing repeated parabolic waveforms[1] are used, either a square or a square root is obtainable, depending on whether the input is a delay and the output a voltage or vice versa. The voltage waveform generated by two integrations of a constant is proportional to the square of the time interval from a reference time; thus, if the voltage output at a given delay is measured, the device produces a square, whereas if the delay at a given voltage is measured, it produces a square root. In this case, over-all feedback is unnecessary. A possible instrumentation of the block diagram uses a parabolic waveform for squaring, with the r- and h-inputs represented by delays (Fig. 6·6); the two output voltages are added, and their sum is compared with the same parabolic sweep, coincidence occurring at a delay proportional to s. In order that the sweep be unchanged in shape by the amplitude comparisons, cathode followers are used for the comparisons. An appropriate change of scale factor must be made in the

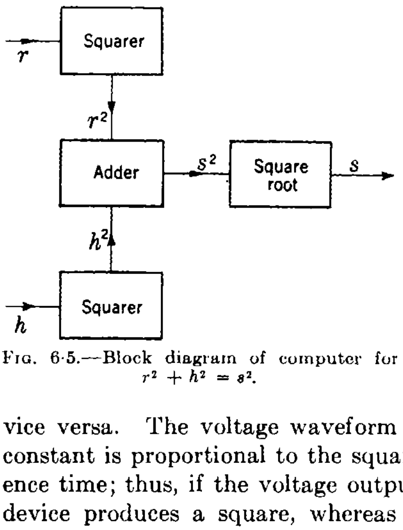


FIG. 6·5.—Block diagram of computer for $r^2 + h^2 = s^2$.

[1] Sec. 5·11; Vol. 19 of this series.

parabola compared with $r^2 + h^2$, for the adding network divides by a factor of 2. Series addition may also be used, in which case two separate parabolic waveforms are generated. A solution using parabolic waveforms in a somewhat different way is given in detail below (Sec. 6·5).

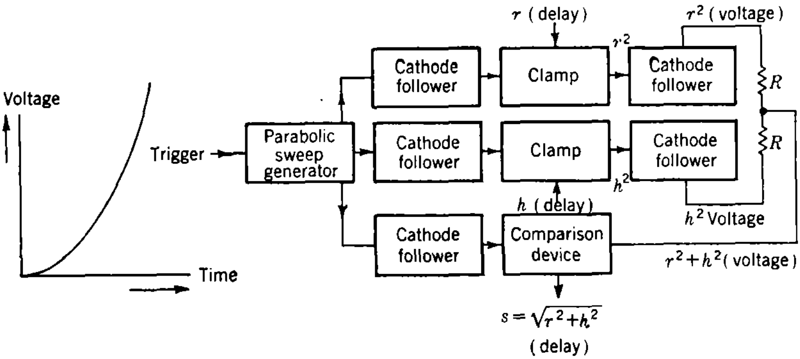


FIG. 6·6.—Hypothetical parabolic-sweep triangle solver.

If the squaring elements used were ganged linear potentiometers (Sec. 5·11), a servomechanism would be required to extract the square root. The squaring blocks would each consist of a pair of ganged potentiometers, and the voltage outputs would be proportional to the input shaft rotations. The two voltages would be added, and the sum applied

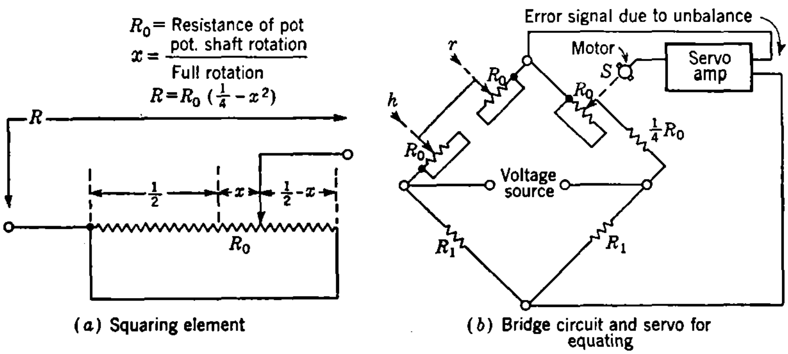


FIG. 6·7.—Triangle solver using resistance squaring and bridge.

to a differential servoamplifier which turned a third pair of ganged potentiometers until their output voltage was the same as this sum. The servo shaft rotation would then measure the length of the hypotenuse.

Method Using Resistance Representation and Bridge for Equating.—It was shown in Sec. 5·11 (and in Sec. 3·5) that a resistance varying parabolically with shaft rotation could be obtained by connecting the ends of a linear potentiometer. This squaring device is shown in Fig. 6·7 together with a triangle solver using three such elements. A Wheatstone bridge is used for equating the resistances of the two arms containing the squaring devices. If the bridge is unbalanced, the error signal changes the resistance representing s^2 in the direction necessary for balance. Either alternating or direct current may be used.

A bridge is an interesting example of a device that solves a relatively complicated implicit function in order to get a simple result. In this case the bridge is solving the equation $y = x$, where y and x represent the resistances in the two upper arms of the bridge. It does this not by producing the function $f(x,y) = y - x$, the simplest implicit function that may be written in this case; it produces the function

$$\frac{E(y - x)}{2(y + x)},$$

where E is the supply voltage, as may be shown by reducing the bridge equations to this special case.

This device requires that component tolerances be held rather closely. The potentiometers representing r^2 and h^2 not only must be linear but must have the same resistance per unit angle to the desired degree of accuracy if the input shafts are to have the same scale of r or h per unit angle. The scale of s per unit angle may be made the same by adjustment of one of the resistors R_1. The input and output dials must be zeroed when the potentiometer arms are in the centers of the respective potentiometers. The resistor marked "$\frac{1}{4}R_0$" must also be adjusted so that when r, h, and s are zero, the bridge is balanced. If all the calibration adjustments are made well enough not to contribute appreciably to the error, the remaining errors will be due chiefly to the matching of resistance per unit angle of the r and h potentiometers and to departures from linearity of all three potentiometers. With 0.3 per cent potentiometer linearity, it should be possible to compute s with errors not exceeding a peak value of 1 per cent of the range of output.

Method Using Electronic Multiplication.—The pulse-length computer developed at Cornell University[1] is applicable to triangle solution. This device solves the equation $ab = cd$ by producing an attenuation a/c and applying it to b to obtain d. In the case of triangle solution the equation $r^2 + h^2 = s^2$ is used in the form

$$h\left(\frac{h}{s - r}\right) - (s + r) = \epsilon$$

with the error signal ϵ amplified and fed back as s, according to the general method of Fig. 6·1. A block diagram of the circuit is shown in Fig.

6·8. The input r and the output s (derived from an amplified error signal) are fed to the blocks "−" and "+", which produce $s - r$ and $s + r$ respectively. The quantity $s - r$ is fed to a multiplying device "×," and the output is made equal to h by an inner feedback loop. The multiplying factor k is thus equal to $h/(s - r)$. In the lower multiplying device, h is multiplied by this same factor, producing $h^2/(s - r)$. This quantity is then compared with $s + r$ by the differential amplifier,[2] and the difference is amplified and fed back as s.

A more detailed circuit diagram is shown in Fig. 6·9. If this figure is compared with Fig. 6·8, it will be seen that the "adder and subtracter" circuit of Fig. 6·9 corresponds to the blocks marked "+" and "−" in

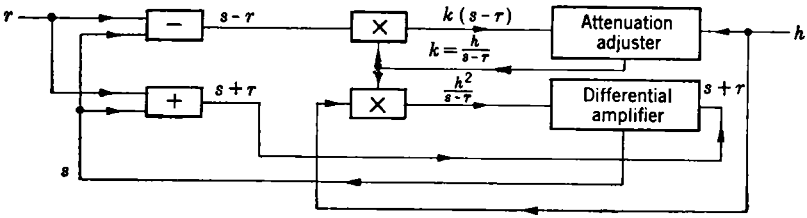


FIG. 6·8.—Block diagram of pulse-length triangle solver.

Fig. 6·8; that the two attenuators correspond to the blocks marked "×"; that the delay multivibrator, gate amplifier, and right-hand differential amplifier comprise the "attenuation adjuster"; and that the left-hand differential amplifier fills the block "diff. amp." The rather complicated attenuator circuit used is necessitated by the fact that the voltage $s - r$ is best taken differentially. This same circuit is used for the lower attenuator, even though the input h is available as a voltage with respect to ground, in order to make the effect of the variable-length gate the same in both attenuators, to balance out the effect of tube drift, and to make level-setting easier. In a number of places the input to a grid is taken from a potentiometer one end of which goes to $B+$ and the other to the input variable. This has the purpose of setting d-c levels, but each such control also affects scale factors.

It is expected that a circuit of this type will solve the equation with errors of the order of 0.5 per cent of output range. The values of resistors are fairly critical, both in bleeders and in plate and cathode circuits where symmetry is desired. In some of the bleeders operating between +450 volts and ground, constancy of output to 0.01 per cent is necessary; this requires good resistors with matched temperature coefficients.

[1] This technique of computation is treated in detail in Sec. 3·12. Circuits for addition and subtraction such as are used in this computation are also treated in Sec. 3·2.

[2] For a more detailed treatment of differential amplifiers see Sec. 3·9.

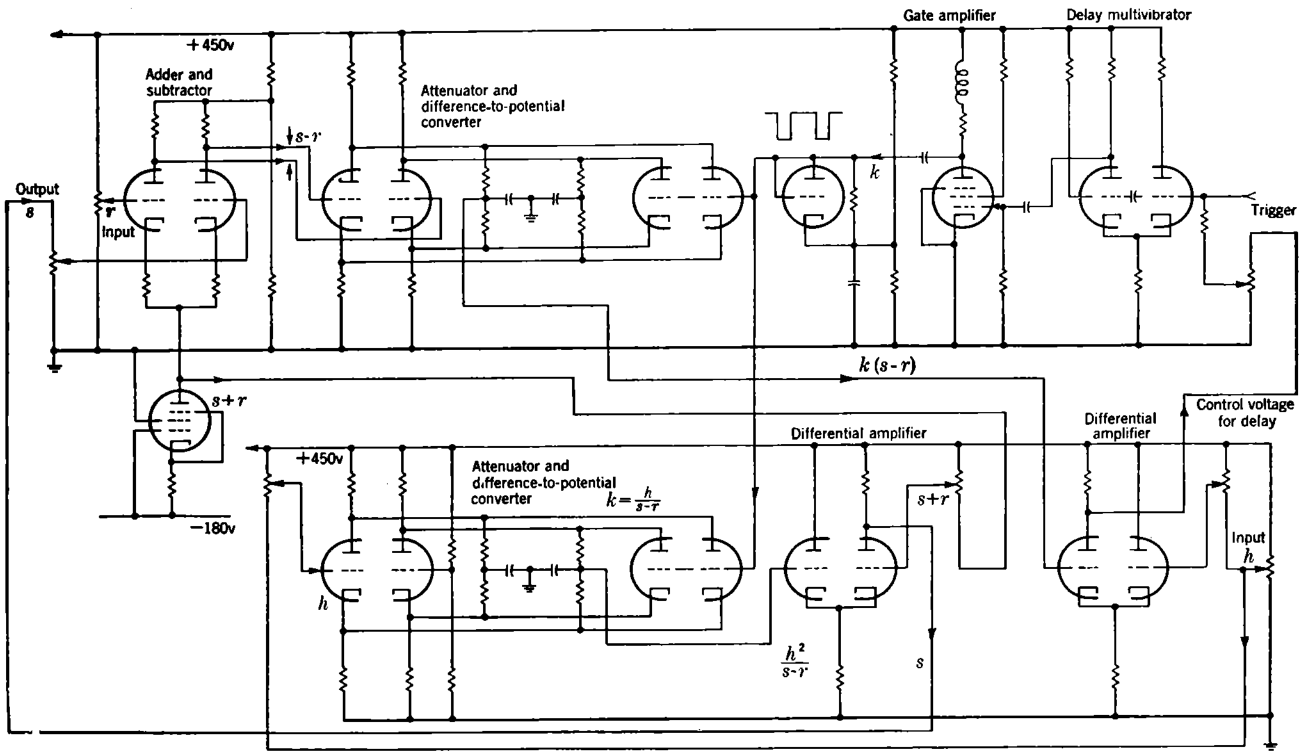


FIG. 6·9.—Solution of $r^2 + h^2 = s^2$ by pulse-length electronic multiplication.

6·5. Curve-fitting Methods for Triangle Solution.—Nonlinear devices constructed by curve fitting[1] may be used for triangle solution. It has been shown in Vol. 19 that a hyperbola may be approximated by an RC-network. This function may be used for triangle solution (Fig. 6·10); for if a waveform is generated whose voltage is proportional to $\sqrt{t^2 - h^2}$ for $t \geqq h$, a voltage measuring r will produce a coincidence pip at a delay[2] $t = \sqrt{h^2 + r^2} = s$. An auxiliary h-delay generator is needed in order to start the hyperbolic waveform at $t = h$. This may be used to provide at the same time a marker on the radar display, so that h may be found by comparison. The accuracy obtainable when three ganged potentiometers are used in the network is about 0.05-mile peak error for $0 < h < 5$ miles and $1 < s < 20$ miles.

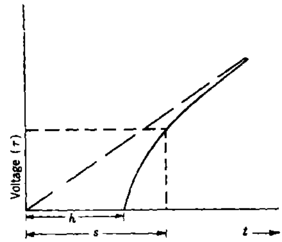


FIG. 6·10.—Hyperbolic waveform.

A device constructed by curve fitting may also be used when a shaft is available whose rotation measures ground range and a voltage proportional to slant range is desired. The operation may be performed with a nonlinear potentiometer if the shaft measures r/h or $\tan\theta$ (Fig. 6·11).

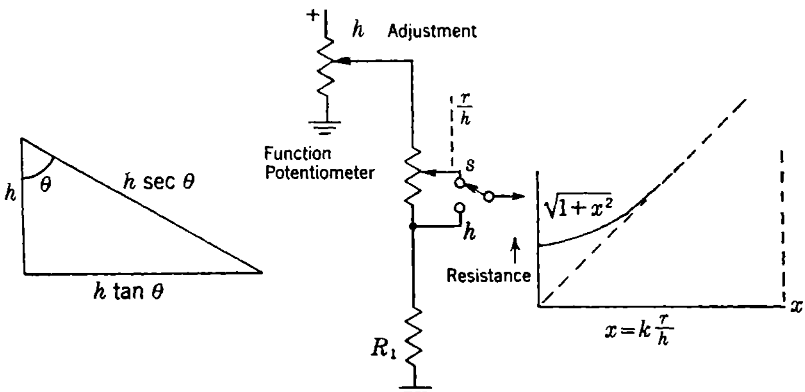


FIG. 6·11.—Hyperbolic potentiometer.

The function produced by the potentiometer may be $\sqrt{1 + x^2}$, where x is shaft rotation; if the applied voltage is h, the output voltage is

$$h\sqrt{1 + \tan^2\theta} = h \sec\theta = s.$$

[1] See Sec. 5·2 for the construction of nonlinear devices by curve fitting.

[2] "A delay s" actually means a delay equal to the time required for a radar pulse to travel a distance s and return, that is, $2s/c$.

The procedure in using it will then be as follows. The output voltage is connected to a delay circuit producing a range mark on the scope. A switch is provided so that the voltage h may be connected directly to the delay circuit; the proper altitude may be set in by adjusting h until the range mark coincides with the first echoes from the ground beneath the aircraft. This same value of h is then used in subsequent operations, and the output of the potentiometer is

$$h \sec \theta = s.$$

A potentiometer for producing this function is discussed in Vol. 20, Sec. 5·2, of this series.

6·6. A Parabolic-waveform Method.—In a method described by F. C. Williams,[1] parabolic waveforms (Vol. 19, Chap. 8) are used for triangle solution by means of the equation

$$\int_0^h kt\,dt - \int_r^s kt\,dt = 0,$$

which is equivalent to $h^2 + r^2 - s^2 = 0$. The circuit generates two sawtooth waveforms of equal and opposite slopes. The positive one is integrated from $t = 0$ to a delay equal to h. The integral of the negative sawtooth waveform, starting at time r, is added algebraically to the first integral; and when this total waveform crosses its original level, an amplitude comparison device produces a pip whose delay measures s. The accuracy of this computation is about ± 0.25 per cent of maximum range.

A circuit diagram is shown in Fig. 6·12. The detailed operation of the circuit[2] may be described with the aid of Fig. 6·13. A positive pulse is applied to the tube V_1 at the terminal A. The front edge of this pulse draws grid current, and the rear edge, which is taken as the zero of time, cuts the tube off for approximately 120 μsec, producing a positive pulse at the plate of V_1 (Fig. 6·13b).

Generation of Sawtooth Waveforms.—In its quiescent state, the plate of V_2 is at about 240 volts because of the d-c feedback from the plate through R_8, R_9, and D_2 to the control grid. During this period D_1 is cut off, its plate being at -5 volts because of R_7. When the plate of V_1 rises, the plate of D_1 is raised, causing the cathode of D_2 to rise and stopping the d-c feedback to the grid of V_2. The current in R_{10} is switched out of D_2 and flows into C_5, and the plate of V_2 runs down at the

[1] F. C. Williams, T. Kilburn, and A. W. Marsh, "A Circuit for the Solution of the Ground Range/Slant Range/Height Triangle in Airborne Radar," TRE Report No. T1844, Apr. 22, 1945.

[2] The following description is taken from a memorandum of F. C. Williams, with slight changes.

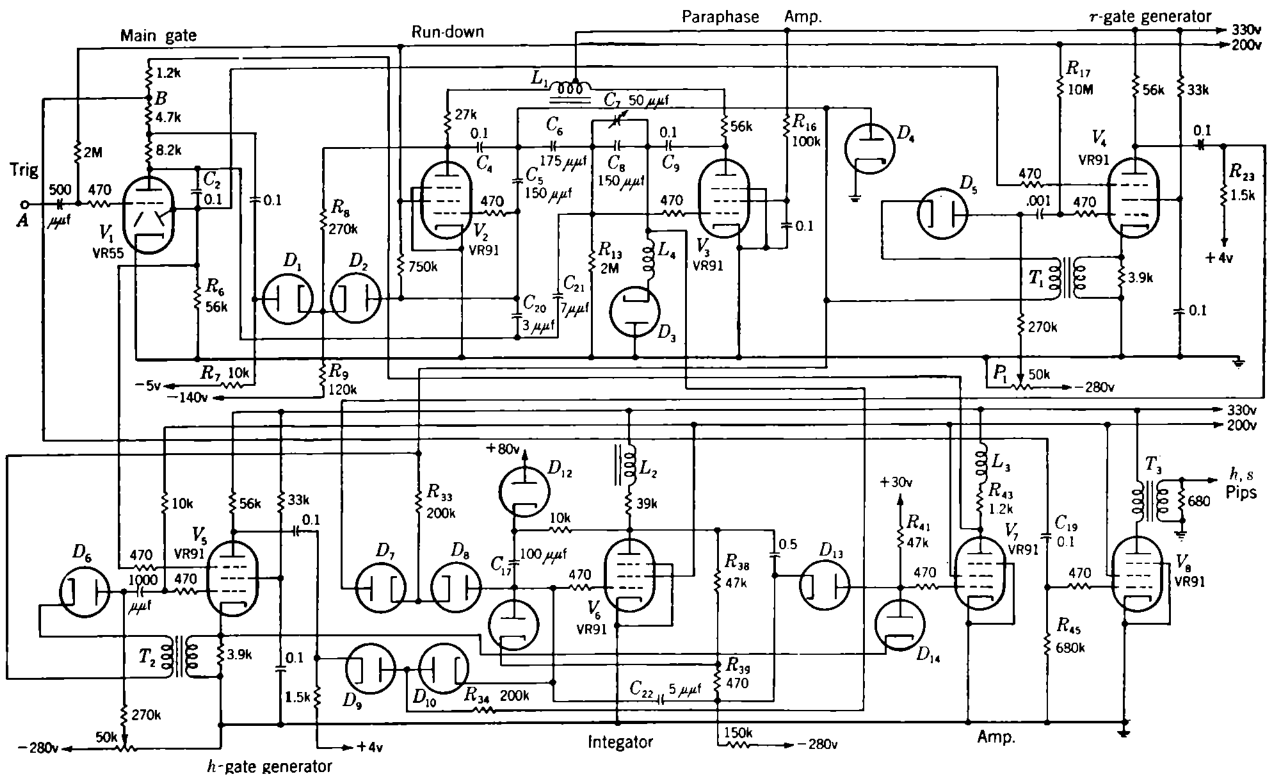


FIG. 6·12.—Parabolic-sweep triangle solver.

rate $200/R_{10}C_5 = 2$ volts/μsec. The inductance L_1 in the plate circuit increases the gain and therefore reduces the distortion of the sawtooth. This rundown continues until the knee of the tube characteristic is reached, that is, until the plate is at about 30 volts above ground. The duration of the sawtooth waveform is therefore $(240 - 30)/2$ μsec = 105 μsec. The plate remains in this condition until the plate of V_1 falls again, at which time the d-c feedback restores it to its original level of 240 volts. The sawtooth is d-c restored[1] to ground by the diode D_4, C_4, and R_{33}. At time $t = 0$ the sawtooth voltage should be equal to the integrator grid voltage, that is, $-1\frac{1}{2}$ volts, so that the current into the integrator V_6 will start from zero. Hence at $t = 0$, a downward step of $1\frac{1}{2}$ volts is produced at the plate of V_2 by applying a positive step to the grid of V_2 from the plate of V_1 via the small condenser C_{20}. The amplitude of the step produced is C_{20}/C_5 times the amplitude of the waveform at the plate of V_1 (see Fig. 6·13c).

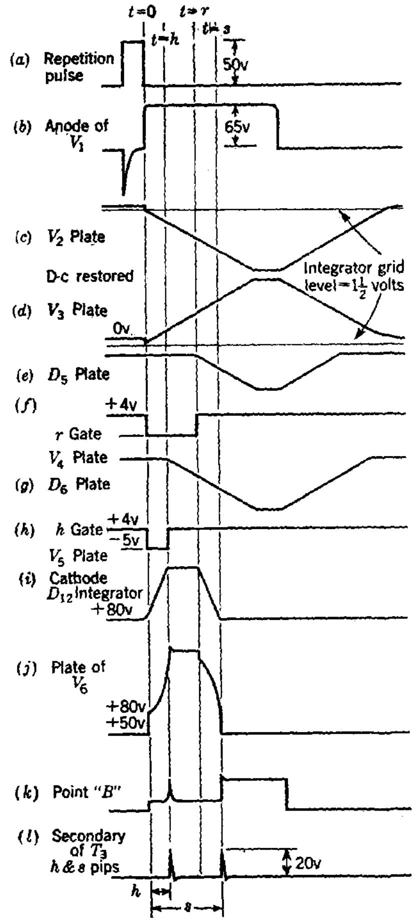


FIG. 6·13.—Timing diagram for parabolic-sweep triangle solver.

In its steady state the paraphase[2] tube has its grid at ground level because of R_{13}, and the space current (and therefore plate current) is limited by the screen voltage, which is determined by R_{16}, so that the plate potential is just above the knee of the characteristic. The rundown sawtooth is applied to the grid through C_6; and if $C_7 + C_8 = C_6$, the plate waveform of V_3 is in paraphase with the plate waveform of V_2. Condenser C_7 is variable so that this condi-

[1] If a waveform is constrained to start from ground level, it is said to be "d-c restored" to ground.

[2] Two waveforms are said to be in *paraphase* if their sum is a constant.

tion may be satisfied. The run-up sawtooth waveform produced is d-c restored to ground by C_9, D_3, and R_{34}, and therefore in the absence of C_{21} the sawtooth wave would have a positive step of $1\frac{1}{2}$ volts at $t = 0$. But since $C_{21} = 2C_{20}$, this positive step of $1\frac{1}{2}$ volts becomes a negative step of $1\frac{1}{2}$ volts, and again the sawtooth wave starts from the potential of the integrator grid (Fig. 6·13*d*). The inductance L_4 allows this negative motion but does not impair the d-c restoration.

Action of the r- and h-switches.—The plate waveform of V_1 is d-c restored to ground by C_2, R_6, and one of the diodes of V_1 and applied to the suppressor grid of V_4. Before the time $t = 0$, the suppressor of V_4 is cut off, and the screen of V_4 is taking all the space current (about 8 ma). The tube is taking grid current through R_{17}, and the plate of D_5 is negative with respect to its cathode by an amount determined by the setting of P_1. When the plate of V_1 rises the suppressor grid of V_4 rises, and 6 ma of plate current flow in V_4, producing a negative edge of 9 volts ($6 \times R_{23}$) at the plate of V_4. This takes the plate of D_7, which was previously at +4 volts, below the level of the integrator control grid, thus opening the *r*-switch, that is, rendering D_8 operative (see Vol. 19, Chap. 3, for switch operation). Current can now flow through R_{33} and D_8 from the run-down sawtooth waveform, and integration commences. The *r*-switch is closed when the run-down sawtooth at the cathode of D_5 reaches the plate potential of D_5; the control grid of V_4 begins to fall, and regenerative feedback through the transformer T_1 turns off the control grid almost instantaneously.[1] The positive edge produced at the plate raises the plate of D_7 and therefore the cathode of D_8 and closes the *r*-switch. Thus V_1 and V_4 define accurately the time interval 0 to *r* and allow the integrator V_6 to integrate the run-down sawtooth during this period only (see Fig. 6·13*b*, *e*, and *f*). The use of V_4 with the run-down sawtooth changes the representation of *r* from voltage (at P_1) to time delay.

The *h*-gate tube V_5 operates similarly, producing the waveform shown in Fig. 6·13*h* in its plate. Before the zero of time the cathode of D_9' is at +4 volts and D_{10} is operative. At $t = 0$, the cathode of D_9 falls to −5 volts cutting off current in D_{10} and closing the *h*-switch. At $t = h$, the cathode of D_9 rises to +4 volts, opening the *h*-switch, and current flows through R_{34} and D_{10} from the run-up sawtooth voltage. Thus the time $t = h$, after which the integrator V_6 is allowed to integrate the run-up sawtooth voltage, is accurately defined by V_5.

Action of Integrator: Production of Parabolic Waveforms.—As the *h*-switch is closed and the *r*-integration begins, the plate of the integrator tube V_6 starts from about 50 volts because of the d-c feedback through

[1] This amplitude-comparison circuit is known as the "multiar" and is discussed in Vol. 19, Sec. 9·14.

R_{38}, R_{39}, R_{40}, and D_{11}. The integrator has two equal leaks, the r-leak R_{33} and the h-leak R_{34}, and two feedback condensers, C_{17} and C_{22}. At the end of the integrator waveform it is essential to produce a sharp edge from which may be derived a slant-range pip; this is the purpose of the diode D_{12}, the resistance R_{36}, and the additional condenser C_{22}. The main integrator waveform produced by C_{17} appears at the cathode of D_{12}, but the plate of V_6 must rise from 50 to 80 volts before any waveform can appear there. The plate is raised to 80 volts by a secondary integration due to C_{22}, which is much smaller than C_{17}. This integration, which is very swift, takes place before and after the main integration, and the final edge cuts off the grid of V_7 via C_{18} and D_{13}, producing the slant-range edge s. The waveforms at the cathode of D_{12} and the plate of V_6 are shown in Fig. 6·13*i* and *j*. (For very small values of r the entire integration is performed by C_{22}.) When the r-switch is opened, negative current flows from the run-down sawtooth through R_{33} and D_8 into C_{22}, so that the plate of V_6 rises parabolically but very swiftly. Hence the cathode of D_{11} is raised, and the d-c feedback removed. When the plate reaches 80 volts, the larger condenser C_{17} takes over and a waveform begins to appear at the cathode of D_{12}. The cathode of D_{12} rises parabolically until the r-switch is closed or until the h-switch is opened. In either case integration stops because the resultant current at the control grid is zero. For if $h > r$, both switches are closed during the time from r to h; if $h < r$, the currents in R_{33} and R_{34} are equal and opposite during the time from h to r, since the slopes of the run-down and run-up sawtooth wave are made equal[1] by the preset condenser C_7. At this point the current flowing through the resistance R_{36} to charge C_{17} is stopped, and the plate of V_6 falls slightly. After this time both the cathode of D_{12} and the plate of V_6 are stationary until the h-switch opens or the r-switch closes. A small step then appears at the plate of V_6 due to the reverse charging current in R_{36} and is followed by a parabolic fall, since the run-up sawtooth voltage is now operating on the integrator. When the cathode of D_{12} reaches 80 volts, it is held by current through D_{12}, and the integration is rapidly completed by C_{22}. The inductance L_2 in the plate of V_6 keeps errors in integration small by reducing the required motion of the control grid.

Production of Markers.—The times h and s are now present in the circuit as edges at the cathode of V_5 and the plate of V_6 respectively. It remains to produce from these edges pips to act as brightness markers on a CRT. The tubes V_7 and V_8 perform this function. The grid of V_7 is normally held at a potential slightly lower than the potential of the integrator grid by D_{13} and R_{41}. When integration commences, the

[1] Actually the zero integration may be made a criterion of calibration; thus the setting of C_7 will also take into account inequalities of R_{33} and R_{34}.

cathode of D_{13} is quickly raised by the plate of V_6 via C_{18}, and V_7 takes grid current through R_{41}. At the same instant the plate current of V_1, which is flowing through R_{43}, is cut off. As a result the point B between R_3 and R_4 rises to a steady level which is maintained until the time $t = h$. At this instant, a negative pip, 5 volts in amplitude and about $\frac{1}{4}$ μsec in width, is produced in the cathode of V_5 and applied to the grid of V_7 via D_{14}. The shape of this pip is maintained in spite of the strays from grid to ground by the comparatively heavy current through the small leak R_{41}, and a positive pip appears in the plate of V_7 and at point B. At the end of the integrator waveform the grid of V_7 is again cut off, via C_{18} and D_{13}, producing a positive step in the plate of V_7 and at B. The sharpness of this edge is enhanced by L_3. The point B returns to its original level when the grid of V_1 is turned on again; the resultant waveform is shown in Fig. 6·13k.

This waveform is d-c restored to ground by C_{19}, R_{45}, and the grid of V_8 and produces the required positive h- and s-pips across a low impedance in the secondary of T_3 (Fig. 6·13l). This d-c restoration ensures that the s-pip occurs at the instant when the integrator tube returns to its initial condition. Transformer T_3 is a differentiating transformer with a 3/1 step-down ratio. The width of these pips is approximately $\frac{1}{2}$ μsec, and their amplitude 20 volts.

6·7. Phase-shift Triangle Solution.—The addition of two alternating voltages that are 90° out of phase may be considered a vector representation of the right triangle in which the magnitude of the sum of the voltages measures the hypotenuse and the phase difference measures one of the acute angles. If the two voltages representing r and h are originally in the same phase, it is necessary to shift the phase of one (usually the h-voltage) by 90°.

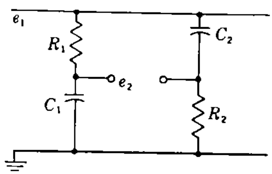


FIG. 6·14.—Network for 90° phase shift.

RC-networks.—The simplest method of doing this approximately is to use a single RC- or CR-network; this, however, has the disadvantage that the phase shift is not quite 90°, and the voltage amplitude is attenuated considerably if the shift is to be nearly 90°. A better method is to use a bridge composed of two such networks, one introducing a phase lead and the other a lag (Fig. 6·14). The ratio of output to input for this combined network is

$$\frac{e_2}{e_1} = \frac{\frac{1}{j\omega C_1}}{R_1 + \frac{1}{j\omega C_1}} - \frac{R_2}{R_2 + \frac{1}{j\omega C_2}} = \frac{1 + \omega^2 R_1 R_2 C_1 C_2}{1 - \omega^2 R_1 R_2 C_1 C_2 + j\omega(R_1 C_1 + R_2 C_2)},$$

and the phase shift is 90° if

$$\omega^2 = \omega_0^2 = \frac{1}{R_1R_2C_1C_2}.$$

The expression may be rewritten

$$\frac{e_2}{e_1} = \frac{1 + \left(\dfrac{\omega}{\omega_0}\right)^2}{1 - \left(\dfrac{\omega}{\omega_0}\right)^2 + j\omega(R_1C_1 + R_2C_2)}. \tag{1}$$

The sensitivity of phase shift to changes in ω is given by

$$\left(\frac{d\phi}{d\omega}\right)_{\omega_0} = \frac{-2}{\dfrac{1}{R_1C_1} + \dfrac{1}{R_2C_2}}.$$

In the case $R_1C_1 = R_2C_2$, the output amplitude is equal to the input amplitude, regardless of frequency. This method of phase shifting has the disadvantage of high output impedance; furthermore, in order to get the output voltage with respect to ground, a transformer or similar device must be used.

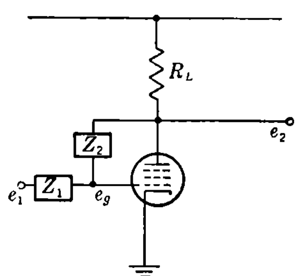


FIG. 6·15.—Single-stage feedback integrator.

Phase Shifting by Feedback.—A method of phase shifting that affords lower output impedance and greater freedom of design uses a feedback integrator and differentiator.[1] The same type of circuit that is used in the parabolic-sweep triangle solver (Sec. 6·6) for integrating linear sawtooth waveforms may be used here for integrating sine waves. A general feedback circuit of this type is shown in Fig. 6·15. The equations of operation may be written for a sine-wave input.

Let e_1 = input voltage,
e_2 = output voltage,
e_g = grid voltage,
i = current flowing from grid to plate through Z_2,
i_L = current in R_L,
Z_1 = impedance between input and grid,
Z_2 = impedance between output and grid,

and assume $i_p = g_me_g$ for a pentode. The derivation can be extended to triodes by letting R_L represent the parallel combination of the load

[1] This method is due to W. G. Proctor. For a detailed treatment of the feedback integrator and differentiator see Secs. 4·1 and 4·7; also Vol. 19, Chap. 18.

resistance and the plate resistance. It is also assumed that no grid current flows. By straightforward circuit analysis it may be shown that the solution for over-all gain is

$$\frac{e_2}{e_1} = -\frac{\dfrac{Z_2}{Z_1}\left(1 - \dfrac{1}{g_m Z_2}\right)}{1 + \dfrac{1}{g_m Z_1} + \dfrac{1}{g_m R_L}\left(1 + \dfrac{Z_2}{Z_1}\right)}. \tag{2}$$

For an integrator, where

$$Z_2 = \frac{1}{j\omega C} \qquad \text{and} \qquad Z_1 = R,$$

this becomes

$$\frac{e_2}{e_1} = \frac{\dfrac{j}{\omega RC}\left(1 - \dfrac{j\omega C}{g_m}\right)}{1 + \dfrac{1}{g_m R} + \dfrac{1}{g_m R_L}\left(1 - \dfrac{j}{\omega RC}\right)} \approx \frac{j}{\omega RC}$$

if $g_m R \gg 1$, $g_m R_L \gg 1$, and $g_m/\omega C \gg 1$. The output amplitude is thus inversely proportional to frequency. In most practical designs the impedances of R and C are large relative to that of R_L in order that the input impedance may be high. Subject to this approximation, the angle by which the phase lag exceeds 90° is

$$\delta \approx \frac{1}{\omega RC g_m R_L} \quad \text{radian.}$$

If $\omega RC = 1$ (unity over-all gain) and the tube is a 6AK5 with $g_m R_L = 100$, $\delta = 0.6°$. If a triode (e.g., a 6C4) had been used, δ would have been about 5°.

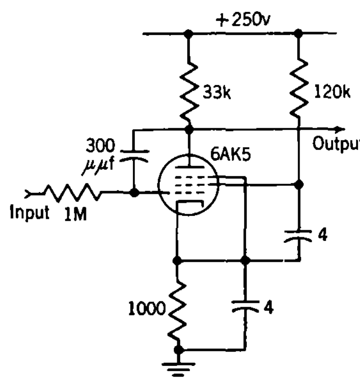


FIG. 6·16.—Integrator circuit.

If the variation of the computing frequency is sufficiently small, an integrator alone can be used as a phase shifter. A possible circuit for use at 500 cps is shown in Fig. 6·16.

If the capacitance and resistance connected to the grid in the previous circuit were interchanged, the circuit would be a differentiator. It would be necessary to insert a blocking condenser in series with the resistance from grid to plate and grid-leak resistor to ground. If, as in Fig. 6·16, the resistance were 1 megohm, a blocking condenser of 0.1 μf would introduce a phase shift of 0.2° at 500 cps. An application of the previous analysis together with Thévenin's theorem shows that the presence of a grid-leak resistor, even of the same order of impedance as Z_1 and Z_2, does not produce an appreciable additional phase shift; its effect is of the order of the $1/g_mZ$ terms rather than the larger $1/g_mR_L$ term.

The output of a differentiator circuit for sine-wave input is then a sinusoid leading the input by 90°; the amplitude response is proportional to frequency. The fact that the amplitude response of a differentiator increases with increasing frequency whereas that of an integrator decreases makes it possible to compensate for the effects of frequency variation by combining the two circuits. The method of combination depends on how the phase-shifted output is to be added to the r-voltage. One method of adding the r- and h-voltages is to use parallel addition with a resistance network (Chap. 3); this results in attenuation as well as addition. Another method, which produces the sum without attenuation, is shown in Fig. 6·17a. The two voltages to be added are applied to the two terminals of a diode; the system assumes an equilibrium state in which conduction occurs at the peaks of $(e_1 - e_2)$. A similar arrangement is possible with a voltage-doubler rectifier, the difference $(e_1 - e_2)$ being applied twice per cycle (Fig. 6·17b). In either case the two phase-shifted voltages from the differentiator and the integrator are added separately to the r-voltage and rectified. If the resulting d-c voltages are then combined by parallel adding, the desired frequency compensation of the output amplitude results. The two alternating s-voltages cannot be added before rectification, for in that case the h-components would cancel one another.

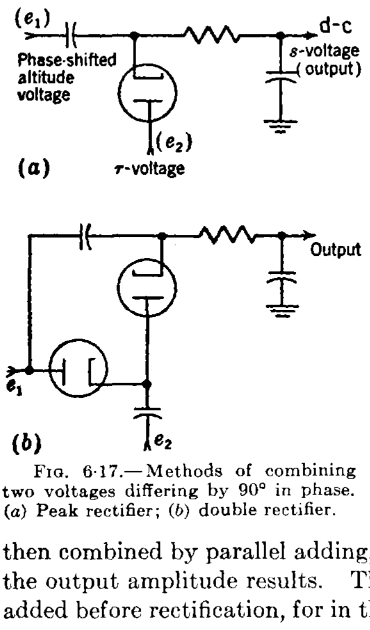


FIG. 6·17.—Methods of combining two voltages differing by 90° in phase. (a) Peak rectifier; (b) double rectifier.

Triangle Solver with Transformer Output.—A third method of combining the r- and h-voltages uses a transformer for addition (Fig. 6·18). This method provides not only the frequency compensation mentioned above but also better compensation for small constant phase differences between the h- and r-voltages that may arise from the manner in which they are obtained in the computer. The two outputs in the circuit of Fig. 6·18 serve this purpose, as will be shown below. Since the tube currents are equal and 180° out of phase, no cathode capacitor is necessary.

The calculations made above for the deviation of its phase shift from 90° and for the amplitude response as a function of frequency no longer hold because the two circuits are no longer independent. The compound circuit may be analyzed in a similar manner.

It will be assumed, as in the above analysis, that the grid-circuit currents are negligible when compared to the tube currents, that the source impedance is zero, and that the phase shifter works into an infinite impedance. In practice, this condition can be more nearly satisfied if a condenser is used for parallel-resonant tuning of the transformer. The tubes will be assumed identical. It will be assumed further that the

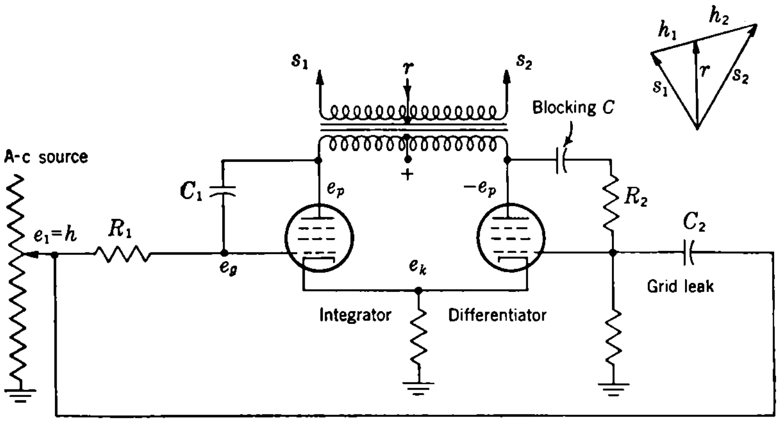


FIG. 6·18.—Phase-shift triangle solver.

transformer is "ideal" to the extent that the two voltages across the halves of the primary are equal and opposite.

Let e_p = plate voltage of integrator,
$-e_p$ = plate voltage of differentiator,
e_1 = input voltage.

It will be assumed that each tube has a voltage gain G. If the plate voltages are equal and opposite, the operation of the integrator can then be expressed by the equation for the grid voltage,

$$\frac{\dfrac{e_1}{j\omega C_1} + e_pR_1}{R_1 + \dfrac{1}{j\omega C_1}} = -\frac{e_p}{G} + e_k,$$

and that of the differentiator by the equation

$$\frac{e_1R_2 - \dfrac{e_p}{j\omega C_2}}{R_2 + \dfrac{1}{j\omega C_2}} = \frac{+e_p}{G} + e_k.$$

By subtraction e_k may be eliminated; the result is

$$\frac{e_p}{e_1} = \frac{1 + \omega^2 R_1 R_2 C_1 C_2}{-1 + \omega^2 R_1 R_2 C_1 C_2 - 2j\omega R_1 C_1 - \dfrac{2\omega^2 C_1 C_2 Z_1 Z_2}{G}}$$

where

$$Z_1 = R_1' + \frac{1}{j\omega C_1},$$

and

$$Z_2 = R_2 + \frac{1}{j\omega C_2}.$$

If the frequency of operation is given by $\omega_0^2 R_1 R_2 C_1 C_2 = 1$, and if G is very large, the expression becomes

$$\frac{e_p}{e_1} = \frac{-\left[1 + \left(\dfrac{\omega}{\omega_0}\right)^2\right]}{1 - \left(\dfrac{\omega}{\omega_0}\right)^2 + 2j\omega R_1 C_1}. \tag{3}$$

This resembles very closely Eq. (1) obtained for the RC phase shifter. At

$$\omega^2 = \omega_0^2 = \frac{1}{R_1 R_2 C_1 C_2},$$

the phase shift is 90°. The difference between this expression and the previous one reflects the fact that the output of the RC-network cannot exceed the input whereas the output of the feedback system can. There is a reversal of sign because a feedback integrator inverts the polarity as well as integrates. The variation of phase shift with frequency is given by

$$\left(\frac{d\phi}{d\omega}\right)_{\omega_0} = R_1 C_1.$$

In the case $\omega_0 R_1 C_1 = 1$ and $G = \infty$, the output amplitude is equal to the input amplitude regardless of frequency; in this case Eq. (3) for the feedback phase shifter has the same form as Eq. (1) for the RC bridge phase shifter. The output impedance of the circuit is roughly equal to the output impedance without feedback divided by the gain of a stage.

If an integrator and differentiator are used without transformer connection in the plate circuit, a considerable decrease in the variation of ϕ with ω may be obtained. This result may be derived with the aid of Eq. (2).

The phase-shifted h-voltage is added[1] to the r-voltage at the transformer secondary as shown in Fig. 6·18. The two solutions for s are

[1] This method is due to J. Lentz.

related as the vector diagram of Fig. 6·18 shows. If the two secondary voltages h_1 and h_2 are equal in amplitude and 180° out of phase, the sum of s_1 and s_2 after they have been converted to direct current represents very nearly the correct solution of the triangle even if the h's are not exactly 90° out of phase with s. This is because the first-order differences of s_1 and s_2 from the true value of s are equal and opposite; the higher-order terms may be found by application of the law of cosines. Thus if the h-input is not in phase with s, or if there is phase shift from primary to secondary of the transformer, the effect will be largely compensated. The compensation is better when transformer output is used than when the outputs of the differentiator and integrator are added separately to r; in the case of separate addition, frequency variation may cause h_1 and h_2 to differ in amplitude, thus making first-order errors from phase shift possible. The transformer circuit, on the other hand, provides both frequency compensation and the balanced output necessary for phase compensation, provided a well-designed transformer is used.

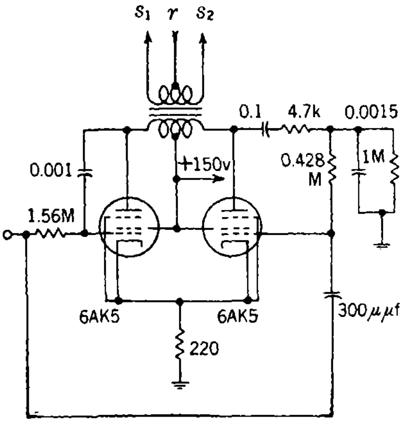


FIG. 6·19.—Integrator-differentiator phase-shifting circuit. (*Courtesy of Bell Telephone Laboratories, New York.*)

Figure 6·19 shows a phase-shifting circuit using a differentiator and integrator and designed to operate at a frequency of 500 cps. The additional network used in connection with the differentiator serves a double purpose: It provides a d-c return for the grid and at the same time permits compensation for the phase-shift error due to the blocking condenser in the feedback circuit. As regards feedback of the plate voltage to the grid, the capacitances 0.1 and 0.0015 μf and the resistances 4.7 k, 1 M, and 0.428 M constitute effectively a Wien bridge. The grid is nearly at ground a-c potential, so that the 1- and 0.428-M resistors are effectively in parallel. If the bridge is balanced, the voltage at the upper end of the 0.428-M resistor is in phase with the plate voltage; this is the desired condition of operation.

Sources of Error.—In a circuit of the type shown, errors may arise from the following sources:

1. *A change in the amplitude of the source* appears as a proportional change in the amplitude of each variable. To eliminate this error,

it is necessary to compensate for these changes; one method of doing this is to use the source as a standard when changing the representation of s.

2. *Frequency change* in the voltage source changes both the amplitude and the phase of the voltage h, as shown above.
3. *Harmonic content in the voltage source* is probably the most serious cause of error in practical applications of a phase-shift triangle solver. It is not difficult, however, to obtain a source having about 0.1 per cent total harmonic content.[1] The error due to harmonic content does not lend itself to simple analysis, but experience has shown that triangle solutions have percentage errors that are of the same order of magnitude as the percentage of total harmonic content.
4. *Potentiometer nonlinearity* obviously causes changes in amplitude of the variables; the resulting error is of the same order of magnitude as the nonlinearity of the potentiometers used.

In the derivations above it is assumed that the impedance of the input voltage source was negligible compared with that of the input resistor (or capacitor). The potentiometer impedance affects the output as though it were in series with the input impedance; it is therefore necessary that this impedance be small. One way of reducing the peak error to about two-thirds its magnitude is to connect an impedance, equal to the input impedance of the triangle solver, between the potentiometer output and the top of the potentiometer, as shown in Fig. 6·20 (see also Sec. 5·4).

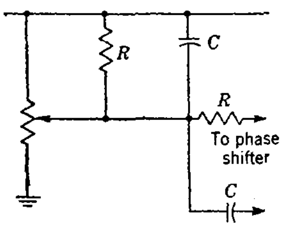


FIG. 6·20.—Method of reducing loading effect on h-potentiometer.

5. The *phase-shift error* introduced by imperfect phase-shift operation is roughly proportional to the square of the phase error in radians if the compensation described above is done accurately.
6. *Change of tube gain* may affect the terms invoving g_m in the expressions for phase-shift and amplitude error. The effect of gain changes is small, however, as long as the gain remains large.
7. *Harmonics introduced by the phase shifter* are usually negligible if the tubes are operated as Class A amplifiers.
8. *D-c drift in the rectifiers* is a "constant" error, that is, it is independent of the voltages being detected. A change in d-c level up to 0.1 volt may be expected in common tubes with aging and with 10 per cent filament voltage changes. This error may be reduced

[1] Chap. 16.

by regular calibration, by the use of a stabilized filament source, or by the selection of a voltage scale such that 0.1 volt represents only a small range. The scale factor is usually determined by letting the maximum voltage available correspond to the desired range of operation.

9. *Unbalance* in the transformer secondary or in the adding circuit may produce first-order errors due to phase shift. If the two voltages r_1 and r_2 are added in such a way that r_1 contributes more to the sum than does r_2, there results a first-order error proportional to the departure from 90° phase shift and to the fractional unbalance. Such a fractional unbalance may result from an asymmetrical center tap of the transformer secondary or from inequality of resistors in the network used to add the d-c voltages. Another sort of error results if the voltages across the two halves of the transformer secondary are not quite in phase with each other; to reduce this it is necessary to wind the transformer symmetrically or to use resistive mixing instead of a transformer to add r and h.

To summarize, the integrator-differentiator triangle solver has the advantage as compared with an RC bridge network of low output impedance, the ratio of input to output impedance being of the order of the gain of a stage. It has the advantage over pulse methods and servo methods of requiring very few parts, including only two pentodes or triodes. A fuller comparison of the methods is difficult unless the desired representations of input and output are stated; for radar applications a rectifier and a time-modulation circuit must be used to convert the output of a phase-shift triangle solver to a delay in order to produce a range marker.

TWO-DIMENSIONAL VECTORS AND TRANSFORMATIONS

When navigation is to be done over relatively short distances, of the order of magnitude of 200 miles, the earth's surface may be considered nearly plane; consequently to a first approximation displacements may be expressed as two-dimensional vectors. Two-dimensional vectors in a plane may be represented by physical quantities resembling vectors, by rectangular components, or by pairs of nonrectangular coordinates that bear some relation to the form in which the navigational data are obtained. Examples of the last sort are polar coordinates, a common form of radar information, and bipolar or two-range coordinates, which make use of distances from two radars of known location. Information obtained in either of these forms may often be used more conveniently if transformed into rectangular coordinates; circuits for performing these transformations and others will be discussed.

6·8. The Mathematical Expression of Transformations.—A correspondence may be established between the points of one plane and the points of another by means of two equations,

$$x = f(u,v),$$
$$y = g(u,v),$$

where the coordinates in one plane are u,v and in the other are x,y. These equations define a transformation of a region of the u,v plane into a corresponding region of the x,y plane. Pairs of relations of this sort enter also into the redescription of a point in a plane in terms of a new set of coordinates, when it has already been described in one set.

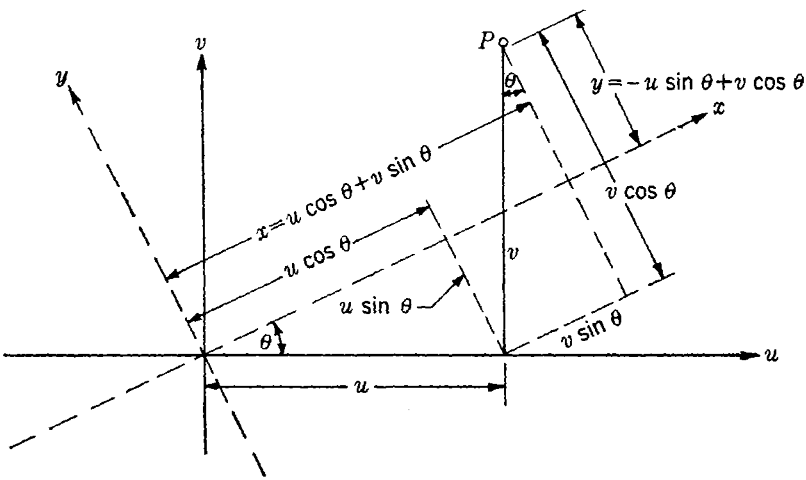


FIG. 6·21.—Rotation of rectangular coordinates.

A special transformation that is of interest is the linear transformation, which in two dimensions is expressed by means of two simultaneous linear equations,

$$x = au + bv,$$
$$y = cu + dv.$$

This corresponds in general to the transformation from one set of oblique coordinates to another. An example of a transformation involving oblique coordinates is given below in connection with the Loran plotting board.

A further specialization may be made to the sort of linear transformation that corresponds to the rotation of rectangular coordinates without change of scale factor. A transformation of this sort may be expressed by the equations

$$
\begin{aligned}
x &= u \cos \theta + v \sin \theta, \\
y &= -u \sin \theta + v \cos \theta,
\end{aligned} \tag{4}
$$

where θ is the angle through which the axes are rotated. The corresponding geometry is shown in Fig. 6·21, relating the coordinates x,y of the point P to the u,v coordinates of the point.

Most of the devices to be described operate in only two dimensions. A general three-dimensional rotation can be accomplished by the use of three two-dimensional rotations of this sort.

6·9. Rotation of Rectangular Coordinates.—To perform the coordinate rotation described by Eqs. (4), devices producing sines and cosines[1] must be used. These may each produce either a single sine function (nonlinear potentiometers) or both sine and cosine together (resolvers, square-card sine potentiometers, phase-shifting condensers). A block

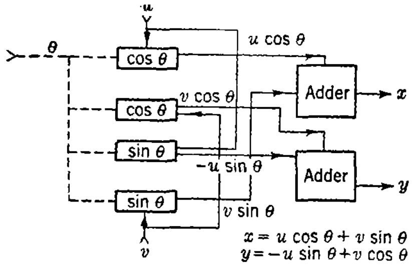


FIG. 6·22.—Rotation of rectangular coordinates.

diagram showing how single sine elements may be connected to rotate coordinates is given in Fig. 6·22. The products $u \cos \theta$ and $v \sin \theta$ are formed and added to give x; similarly, $-u \sin \theta$ is added to $v \cos \theta$ to give y. The accuracy of such a transformation is limited by the accuracy of the components used.

If a resolver is used to rotate rectangular coordinates,[2] either it may be considered to be solving the equations of transformation by producing sines and cosines and adding them, or it may be considered a physical model of the transformation in which the inputs are the projections of the magnetic field on the stator axes and the outputs its projections on the rotor axes. A schematic diagram of the rotation of coordinates by a resolver is shown in Fig. 6·23. If square-card sine potentiometers are to be used, two are necessary, for each produces one sine and one cosine function.

When a-c amplitude and phase represent a vector, coordinate rotation corresponds simply to changing the phase angle. A phase shift varying

[1] For a detailed treatment of sine-cosine devices, see Secs. 5·6 and 5·9.

[2] See also Miller and Weisz, "Coordinate Transformation Circuits Using Resolvers," NDRC Report No. 14-228, June 1, 1944.

linearly with shaft rotation may be obtained by the use of a phase-shifting condenser or a resolver. If the input signal is put on one input terminal and the same signal phase-shifted through 90° on the other input terminal, the rotor will pick up a signal having amplitude proportional to the input and phase shift varying linearly with the angular position of the rotor. Three-phase devices may also be used. The accuracy obtainable in this operation with phase-shifting condensers is about 1°; with resolvers 0.1° (see Vol. 20, Chaps. 12 and 13).

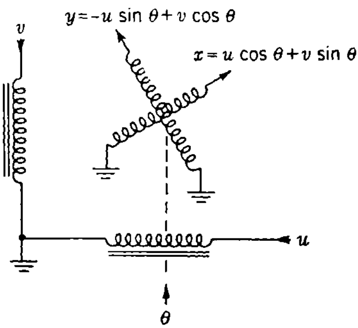


FIG. 6·23.—Rotation of coordinates by a resolver.

6·10. Polar to Rectangular Transformations.—Transformation from polar to rectangular coordinates is simply the production of a sine and a cosine. It may therefore be done as a special case of coordinate rotation, by any of the methods mentioned in Sec. 6·9. If the radial-coordinate input is considered a vector that lies along one of the rectangular axes and is to be rotated through an angle θ (the other polar coordinate), the operation performed is the production of a sine and a cosine. A block diagram of this special case and a diagram showing the use of a resolver to convert polar to rectangular coordinates are shown in Fig. 6·24a and b respectively. The input to the resolver in this case is to the rotor and the outputs from the stators, but the reverse arrangement might equally well be used. In practice it is more common to let the rotor be the input because there are a number of "1-to-2-phase" devices that have only one rotor winding but two stator windings.

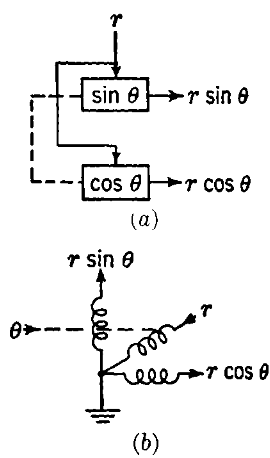


FIG. 6·24.—Polar-to-rectangular transformations.

6·11. Rectangular to Polar Transformations.—The conversion of rectangular to polar coordinates involves somewhat more complication. The equations of transformation, written in explicit form, are

$$r = \sqrt{x^2 + y^2},$$
$$\theta = \tan^{-1}\frac{y}{x}.$$

The operations of squaring, extraction of the square root, division, and production of the inverse tangent are not simple. Implicit solutions are more convenient in this case. The equations of transformation may be expressed in implicit form as

$$x = r \cos \theta,$$
$$y = r \sin \theta,$$

or again as

$$-x \sin \theta + y \cos \theta = 0,$$
$$x \cos \theta + y \sin \theta = r.$$

In the latter pair of equations, the first equation is an implicit one for θ in terms of x and y; the second is an explicit expression for r in terms of x,y, and θ, which can be computed in a straightforward way once the first has been solved. A block diagram showing how this transformation can be performed with single-sine elements is shown in Fig. 6·25a. The function $(-x \sin \theta + y \cos \theta)$ is computed by two of the sine or cosine blocks, and this function is fed to a servoamplifier that drives the θ-shaft until the function reaches zero. When this loop reaches equilibrium, the proper shaft rotation θ will be fed into the other sine-cosine blocks, producing r explicitly.

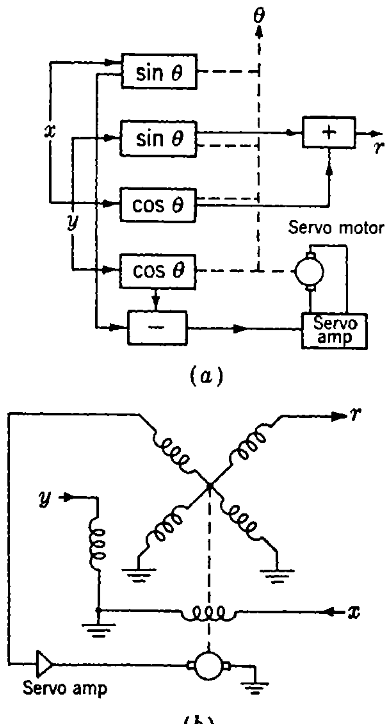


FIG. 6·25.—Rectangular-to-polar transformations.

The implicit solution of the θ-equation is similar to the type discussed in Sec. 6·2 and shown in Fig. 6·1; in both cases a combination of operations generates a function that is desired to be zero, and this function is fed back in such a way as to reduce its value. There is an essential difference between the two methods, however. In the circuit of Fig. 6·1 the function was fed back as one of the input variables, whereas in this case, roughly speaking, it is fed back as a rate of change of one of the input variables θ. Thus the system of Fig. 6·1 has an inherent error resulting from the fact that at equilibrium the function $f(x,y)$ has a small residual value measured by y/A; on the other hand, systems of the type used here, in which the feedback determines the *rate of change* of a variable, will have no such inherent position error at equilibrium, for the error signal is $f(x,y) \approx (dy/dt)/A$, and this is zero because at equilibrium $dy/dt = 0$.

Resolvers.—Figure 6·25*b* shows how this transformation may be performed with a resolver.[1] It will be seen that the operation is the same if it is considered that one rotor winding picks up a voltage

$$-x \sin \theta + y \cos \theta$$

and the other $x \cos \theta + y \sin \theta$. The servoamplifier has the same function as before, and r may again be considered to be found explicitly. Perhaps a simpler physical explanation of what happens is that the servoamplifier causes one rotor to be oriented perpendicular to the magnetic field that is the resultant of the two fields produced by the stators; the other rotor winding, parallel to the field, picks up the resultant voltage. With the Arma resolver (No. 213044) or Bendix resolver (XD-759542) such a transformation can be done with a peak error of about $\pm 0.1°$ in an angle and ± 0.06 per cent of maximum output in output a-c amplitude.

Sine Potentiometers.—The transformation from rectangular to polar coordinates can also be made with a pair of ganged square-card sine potentiometers. A method of this sort may be used with d-c voltage representation. A diagram of this method is shown in Fig. 6·26. Each sine potentiometer performs the operations corresponding to two of the sine or cosine blocks in Fig. 6·25. The principal problem in the design of such a system results from the fact that d-c voltages representing the various quantities involved do not have ground potential as a reference level unless a push-pull supply is available for the sine potentiometer (see Chap. 5). Also, in order to use the sine potentiometers with greatest accuracy, the outputs should be taken push-pull rather than single-ended. One possible approach is to produce balanced positive and negative voltages at the inputs, thereby centering the outputs at ground. This, however, necessitates two voltage-inverting circuits, each of which may require six triode sections (Chap. 3), whereas the method shown requires only one inverting circuit.

The circuit shown provides for cancellation of the voltages Ax and By which are added to the sine and cosine components because the voltage supplies for the sine potentiometers are not symmetrical. The servoamplifier that orients the shafts of the potentiometers must solve the equation

$$x \sin \theta - y \cos \theta = 0.$$

[1] A sample design of a resolver servomechanism of this type is given in Sec. 11·4.

It does this by virtue of the fact that the "constant" terms Ax and By are fed equally to the two inputs of the differential amplifier. The equation actually solved is

$$-\frac{x}{2}\sin\theta + Ax + \frac{y}{2}\cos\theta + By = \frac{x}{2}\sin\theta + Ax - \frac{y}{2}\cos\theta + By;$$

but this is equivalent to

$$x\sin\theta - y\cos\theta = 0.$$

The terms Ax and By are removed from the output voltage by a similar cancellation. The same subtraction that makes possible the push-pull

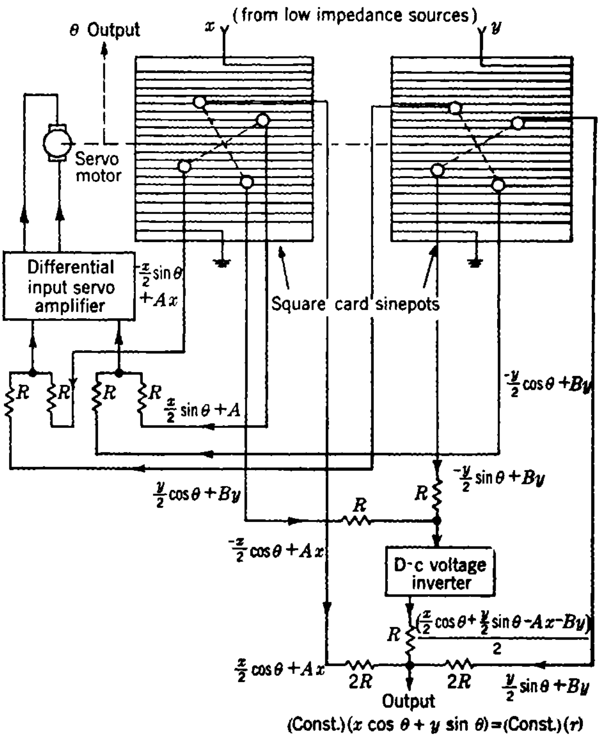


FIG. 6·26.—Rectangular-to-polar transformation using direct voltage.

operation of the output is also used to remove the "constant" voltages. Addition before inversion makes it necessary to use only one d-c voltage-inverting circuit.

There is a loading error that results from connecting the adding resistors R directly to the sine potentiometers. If a sine potentiometer has 20-k resistance and the mixing resistors are of the order of 1 megohm, the error from this source will be of the order of 0.3 per cent of maximum output (see Sec. 5·7). An inconvenience of this circuit (and of all precise adding circuits using resistance networks) is that the resistors require calibration if maximum accuracy is to be obtained. This requires five calibration potentiometers in series with resistors. Since the sine potentiometers may have ± 0.25 per cent peak errors, the over-all peak error of the computation may be as great as 1 per cent.

The Loaded-potentiometer Resolver.—There is another method of transforming rectangular to polar coordinates that requires no special parts like resolvers or sine potentiometers but operates over only a limited range of angle ($\pm 50°$, for example) depending on the degree of approximation desired. This is the "loaded-potentiometer resolver."[1]

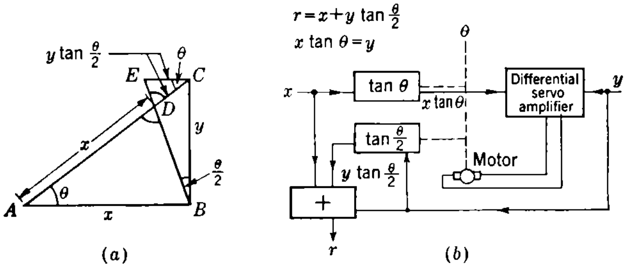


FIG. 6·27.—Loaded-potentiometer triangle solver.

This device makes use of the fact that a tangent function can be approximated by the use of a loaded potentiometer (Sec. 5·4). The equations of transformation used in this case are

$$x \tan \theta = y,$$
$$r = x + y \tan \frac{\theta}{2}.$$

This again is a case of an implicit solution for θ followed by an explicit solution for r. These equations are equivalent to the equations $x = r \cos \theta$, $y = r \sin \theta$; the equivalence is shown geometrically in Fig. 6·27*a*. By laying off a distance x along the hypotenuse, the isosceles triangle ABD is formed; if EC is constructed parallel to AB, the triangle CDE is similar to ABD and therefore isosceles; and $CE = CD$. The angle $EBC = \theta/2$; therefore $CD = CE = y \tan \theta/2$ and $r = x + y \tan \theta/2$.

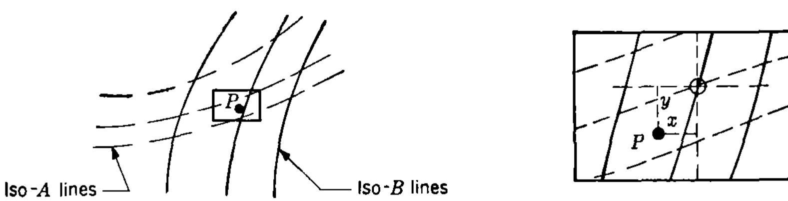
Since two equations of transformation involving no trigonometric functions other than tangents are given and a means of approximating the tangent function is available (Sec. 5·4), a block diagram of a device

[1] G. D. Schott, "Loaded Potentiometer Triangle Solver," RL Group Report No. 63, May 31, 1944. The method is due to J. W. Gray.

for making the transformation may be drawn. This is shown in Fig. 6·27*b*. The function $x \tan \theta$ is produced, and the difference $x \tan \theta - y$ actuates a servoamplifier that turns a motor orienting the θ-shaft; the value of θ computed there is then used to produce $y \tan \theta/2$; this is added to x to give r. The error observed over a range of $\pm 50°$ is approximately 0.5 per cent in r and $\pm 0.2°$ in θ, due chiefly to potentiometer nonlinearity and inaccuracy of calibration.

The transformation from rectangular to polar coordinates may also be done if a phase-shift triangle solver (Sec. 6·7) is used to produce both phase and amplitude outputs. The phase shift may be converted to rotation as shown in Chap. 3.

6·12. Special Coordinate Transformations. *Oblique Coordinates: The Loran Plotting Board.*—In the Loran navigational system (Vol. 4), an aircraft may find its position by measuring the differences in time



(*a*) Intersecting hyperbolas (*b*) Expanded region

FIG. 6·28.—Loran geometry.

delays of synchronized pulses from two pairs of transmitter stations. The time difference (i.e., the difference in range) corresponding to each pair of stations locates the aircraft on one of a family of hyperbolas, of which the two stations are the foci. The measurement of two such time differences, one from each pair of stations, locates the aircraft at a point that is the intersection of two such hyperbolas. In practice, the transmitters may be separated by several hundred miles. In some cases it is desirable for the navigator to know his position in some region whose dimensions are small compared with the separation of the transmitters. In this case the approximation may be made that the coordinates may be transformed linearly in that region (Fig. 6·28).

It has been desired to plot automatically the coordinates of the plane on a board, as determined by the Loran information. Subject to the assumption of straight-line oblique coordinates, there are at least two ways in which such a plotting board may be built so that it can be adjusted to provide for the many different angles of intersection which may be encountered. One such method is to let the plotting point be moved by two mechanical motions whose angles relative to one another are adjustable.[1] A diagram of such a device is shown in Fig. 6·29. In the device shown, one of the coordinates of the board must be parallel to each of the families of hyperbolas in the region considered.

If it is desired that the two mechanical motions of the plotting point be N-S and E-W (rectangular coordinates), this can be done by means of a transformation[2] between the Loran delay information and the information at the board. The Loran time delays may be converted to proportional d-c voltages by the use of delay circuits set manually to produce the same delay as the Loran. These d-c voltages then measure incremental distances in an approximate system of linear oblique coordinates. The procedure is to transform the d-c voltages to express position in N-S and E-W coordinates measured from a predetermined origin in the region considered.

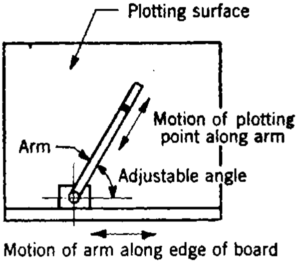


FIG. 6·29.—Plotting board with two mechanical motions at adjustable angle.

The relation of the rectangular and the oblique coordinates is shown in Fig. 6·30. The family of hyperbolas running most nearly east-west will be called A-hyperbolas or iso-A lines, and the others iso-B lines. The voltage corresponding to the A-delay is represented by E_A, and E_B is the voltage corresponding to the B-delay. If E_A varies alone, the plotting point must move on an iso-B line, at an angle β from north. Let a represent the distance of movement *along an iso-B line* required for each volt of E_A. Then if E_A increases by ΔE_A volts and B is constant, the northward movement of the plotting point is $\Delta y = a \cos \beta \, \Delta E_A$ and the eastward movement is $\Delta x = a \sin \beta \, \Delta E_A$. If E_B varies alone, the stylus moves in an iso-A line, at an angle α from east. Let b represent distance per volt of E_B *along an iso-A line*. Then $\Delta x = b \cos \alpha \, \Delta E_B$, and $\Delta y = b \sin \alpha \, \Delta E_B$. If both E_A and E_B change, the increments of x and y give the equations of transformation

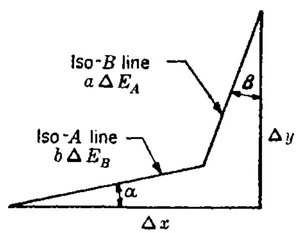


FIG. 6·30.—Relation of rectangular to oblique coordinates.

$$\Delta y = a \cos \beta \, \Delta E_A + b \sin \alpha \, \Delta E_B,$$
$$\Delta x = a \sin \beta \, \Delta E_A + b \cos \alpha \, \Delta E_B.$$

[1] This method was used in a plotting board designed in Division 11 of the Radiation Laboratory.

[2] This method is due to J. W. Gray.

The signs of the various terms depend also on the sign convention adopted for the positive directions of A and B.

A schematic diagram of a computer for doing this is shown in Fig. 6·31, and photographs of a plotting board constructed for this purpose in Fig. 6·32. The instrumentation is as follows. Voltages measuring ΔA and ΔB are obtained by comparing the outputs of time-modulation circuits with the Loran delays. A voltage varying linearly with ΔA and ΔB is obtained by a resistive mixing network; the coefficients of ΔA and ΔB are adjusted by using potentiometers as shown in the diagram. Since for a given mission the values of a, b, α, and β are constant, they may be set in advance. The final step of converting voltage to

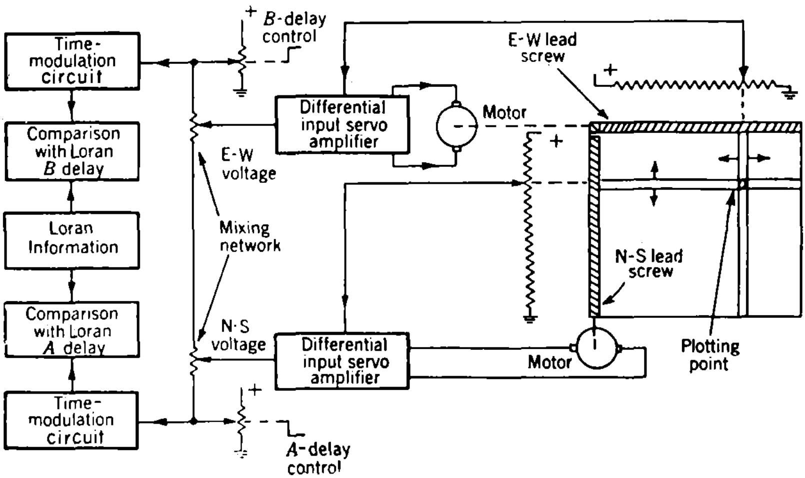
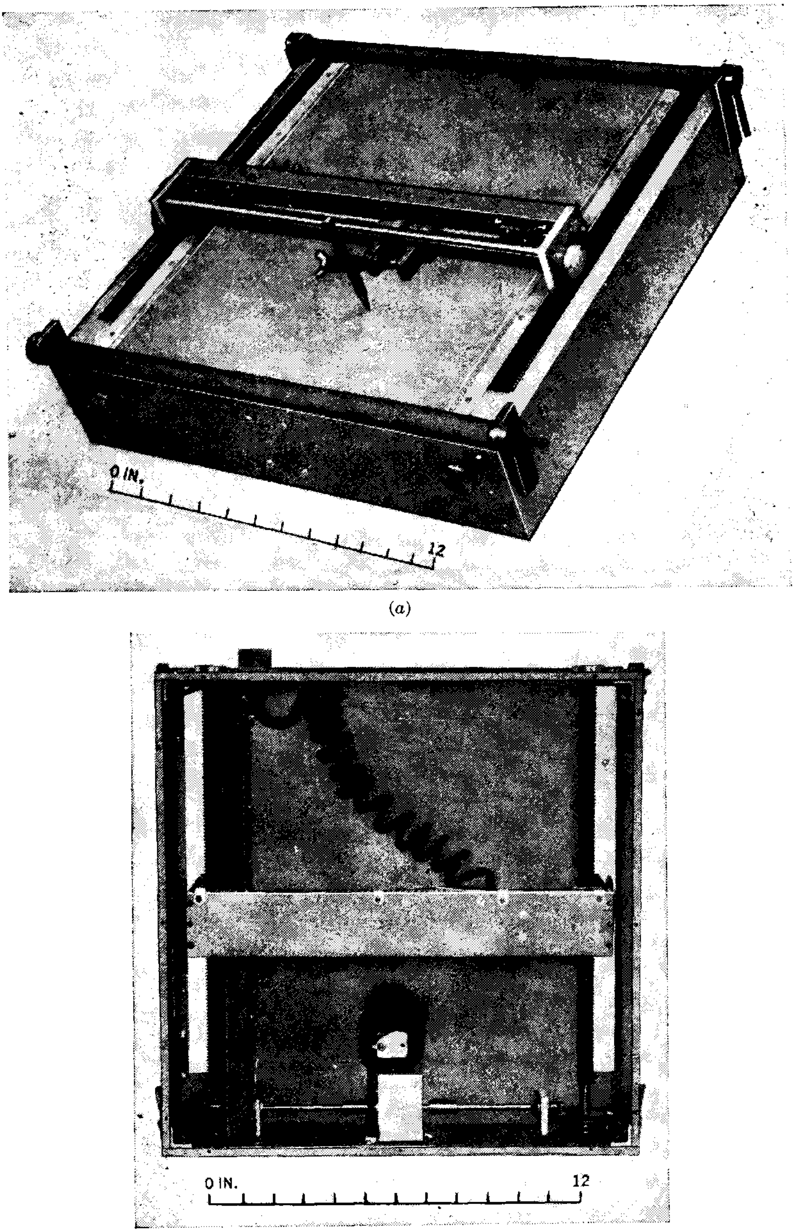


FIG. 6·31.—Schematic diagram of Loran plotting board.

displacement of the plotting point is done by means of linear-card potentiometers and servos operating lead screws. The E-W lead screw and potentiometer are actually moved across the board by two N-S lead screws, as shown in Fig. 6·32. The plotting point moves in each coordinate until the voltage picked up by the potentiometer arm is equal to the transformed voltage in that coordinate. The adjustment of the potentiometers may be checked by observing whether or not the plotting point moves properly when each time delay is varied. Switches must also be provided to allow for different orientations of the hyperbolas.

Bipolar Coordinates: A Transformation Using Squaring Devices.—The Loran plotting board, which has just been described, is an instance of the use of radar range data for locating a point in two dimensions. Another problem of similar nature arises when the ranges of an aircraft from two fixed stations on the ground are known and it is desired to



(b)

FIG. 6·32.—Loran plotting board. (a) Top view; (b) bottom view.

express the position of the aircraft in rectangular coordinates. This constitutes a sort of bipolar coordinate system, but not an orthogonal system. A pair of rectangular axes convenient for computation has its origin at one of the fixed stations and its y-axis passing through both stations. In this case the equations of transformation are no longer linear. They are

$$y = \frac{l}{2} + \frac{m^2 - n^2}{2l}$$
$$x^2 = m^2 - y^2,$$

where x, y = the rectangular coordinates as described above,
l = distance between the two fixed points,
m = range from fixed point at origin,
n = range from other fixed point.

The geometry is shown in Fig. 6·33a. A computer for solving this problem was designed by H. S. Sack[1] and uses methods similar to those used

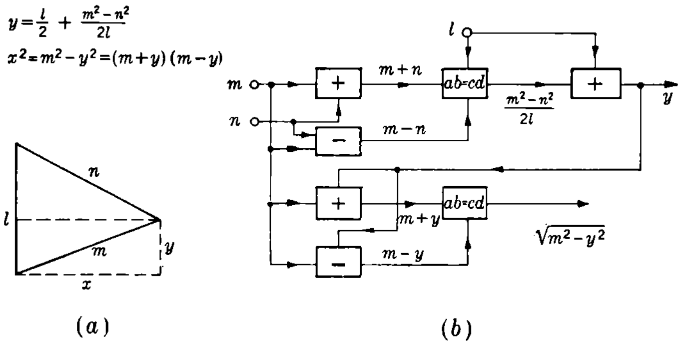


FIG. 6·33.—Bipolar to rectangular transformation. (a) Geometry; (b) block diagram of computer.

in the electronic triangle solver of Sec. 6·3. A block diagram of this device is shown in Fig. 6·33b. In the computation of y the operation of squaring is carried out by a device that solves the equation (Chap. 3)

$$\frac{m + n}{2l} = \frac{u}{m - n}$$

or

$$m^2 - n^2 = 2lu.$$

(Here u is an intermediate variable in the computation of y.) The quantity $l/2$ is then added to u. This is done by the four blocks in the upper part of the diagram. Similarly the lower three blocks compute x

[1] H. S. Sack, "Report on Computers Involving Squares and Square Roots," Cornell University, Oct. 19, 1943. This work was done under an OSRD contract.

as a function of m and y, by solving the equation

$$\frac{m+y}{x} = \frac{x}{m-y}.$$

A schematic circuit diagram for this computer (from the report referred to above) is shown in Fig. 6·34. Differential amplifiers with

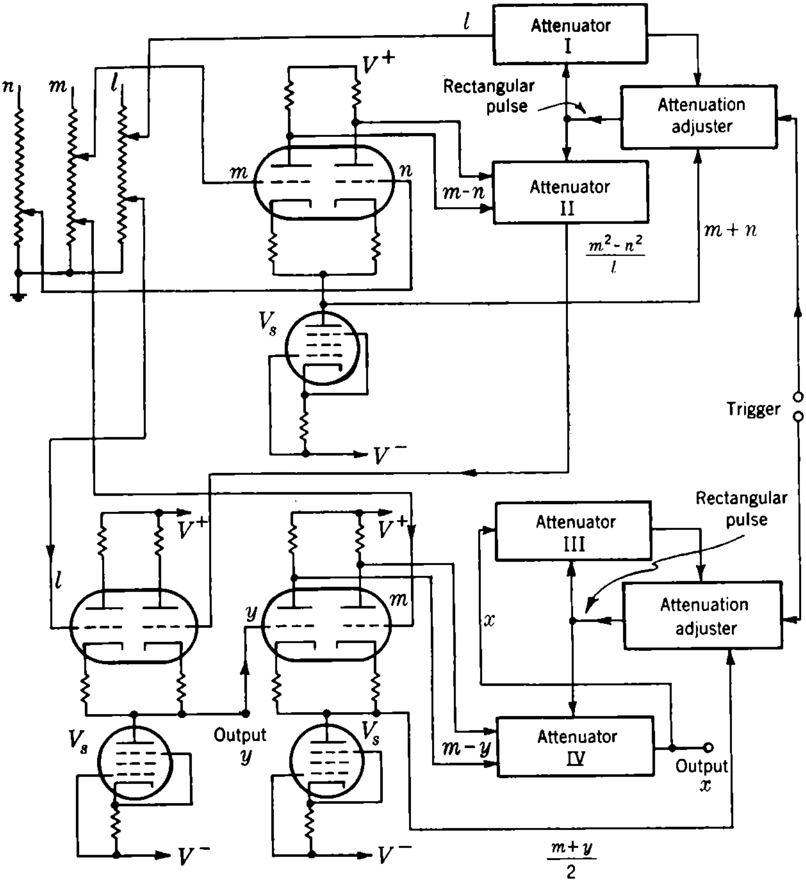


FIG. 6·34.—Computer for $y = (l/2) + (m^2 - n^2)/2l$; $x^2 = m^2 - y^2$.

constant-current tubes in the cathode circuits are used for addition and subtraction. Each of these requires two envelopes: a pentode and a double triode. Each "attenuator" block is a differential amplifier which is made inoperative for part of the repetition interval by a variable-length rectangular pulse from the attenuation adjuster. The feedback loop formed by the attenuation adjuster and the upper attenuator assumes an equilibrium condition in which the output of the upper attenuator is equal to the other input to the attenuation adjuster. The rectangular pulse then causes the attenuator to multiply by a factor $(m + n)/l$, in the case of the y-computation. The same pulse then determines the attenuation of the lower attenuator block, the output of which will be $(m - n)(m + n)/l$.

The circuit as shown requires about 28 tube envelopes, not including VR tubes or the trigger circuit. This number may be reduced if lower precision is required or if different methods of attenuation and addition are used. The error in y may be 0.25 per cent, that of x probably somewhat greater.

CHAPTER 7

EXAMPLES OF COMPUTER DESIGN

BY D. MACRAE, JR., I. A. GREENWOOD, JR., AND W. ROTH[1]

7·1. Introduction.—Preceding chapters have covered methods of design, techniques, and devices for use in the creation of computers. It is the object of this chapter to present two typical computers to illustrate the application of this information. The computers that will be discussed are (1) part of an airborne navigation computer and (2) a radar trainer computer solving the problem of synthesizing the position of an aircraft in polar coordinates relative to a moving ship.

The first computation to be discussed is one that solves part of the problem of aircraft navigation using radar. The computer of which it is a part is a ground-position indicator[2] which integrates the airspeed of an aircraft and the wind to provide a continuous indication of position; it makes use of radar to obtain an accurate indication of the position of the aircraft with respect to identifiable objects on the radar screen and to find wind. The computation to be described has to do with the production of markers on a radar screen, corresponding to two shaft rotations that represent the rectangular coordinates of a point with respect to the aircraft; this constitutes only a portion of the entire navigational computer. The treatment will be detailed in order to show some of the practical considerations that enter into a design. The discussion of this computer will follow the general design procedure given in Sec. 2·1; at each step in the design the material of Sec. 2·1 is summarized, followed by a discussion of the corresponding steps in the development of the navigational computer.

At the time of termination of Radiation Laboratory technical development work, the design of this computer was not finished, though most of the circuits in it were completed or nearly completed. It is expected that the development of this computer will be carried on by organizations other than the Radiation Laboratory.

[1] Sections 7·1 to 7·7, inclusive, are by D. MacRae, Jr., and I. A. Greenwood, Jr.; Secs. 7·8 to 7·11, inclusive, are by W. Roth and I. A. Greenwood, Jr.

[2] B. Chance, "The Interconnection of Dead Reckoning and Radar Data for Precision Navigation and Prediction," *Jour. Franklin Inst.*, **242**, pp. 355–372; W. J. Tull, N. W. MacLean, "GPI—An Automatic Navigational Computer," *Jour. Franklin Inst.*, **242**, pp. 373–398, November 1946.

The spherical-coordinate integrator, the second computer described, was used extensively in radar training equipment. In the discussion of this equipment, the emphasis is on the description of the device rather than on the process of its design. In several instances different circuits are used in the two computers to perform the same operations. This is due mainly to different accuracy and weight requirements but also partly to the fact that different personnel were involved in the two designs.

NAVIGATIONAL COMPUTER

7·2. Preliminary Information. *Procedure of Sec.* 2·1.—It was stated in Sec. 2·1 that the first step in the design of a computer is the determination of what is to be computed and the interpretation of functional needs in terms of technical specifications and basic equations or empirical relations. At this point the designer also makes tentative over-all block diagrams, using his knowledge of computing elements. Judgment and ingenuity are particularly required in this step. The characteristics of the basic data, the factors that limit the designer's choices, and the operating conditions must be determined. Desired controls, displays, and outputs must be specified, and acceptable alternatives listed. At this stage in the design it is appropriate to start discussing with component specialists the possibilities of getting some of the scarce components that might be useful in the design.

The GPI Navigational Computer.—A ground-position indicator (GPI), designed at the Radiation Laboratory, makes use of radar information both for locating the aircraft and for determining wind. Both these operations use a group of circuits that transform the mechanical information in the computer into markers whose intersection constitutes an index of position on the radar display. These are the circuits which are discussed here in detail.

The air speed is resolved into north-south and east-west components, and these are added to the respective components of wind, giving the two resultant components of the velocity of the aircraft relative to the earth. These are integrated by means of electromechanical integrators of the type described in Sec. 4·9. The outputs of the two integrators are shaft rotations representing the position of the aircraft and turning at a rate determined by the values of air speed and wind which are integrated. The relation of these coordinates (aircraft position) to the other coordinates involved is shown in Fig. 7·1. These shaft rotations can be subtracted from another pair of adjustable shaft rotations that represent the coordinates of a reference point identifiable on the radar display, and the differences represent the coordinates of the reference point relative to the aircraft. These differences of coordinates may be used in locating the aircraft with respect to identifiable echoes on the radar screen and in finding wind. Let the index be set to coincide with an identifiable echo on the screen by first setting in the known coordinates of the point and then bringing the index to coincidence by rotating the shaft measuring aircraft position. When this is done, a *fix* is said to have been taken. The readings of the aircraft position shafts will represent the actual location of the aircraft.

The same index can also be used for finding wind. Since the coordinate differences from which the index is formed are changing with a velocity corresponding to the vector sum of air speed and wind, the index will remain on an echo once it has been set to coincidence with it provided all velocities have been correctly entered. The index and echo move across the screen together as the aircraft flies past the identified point. If there is an error in the wind entered into the integrator, the index will drift off the echo. This information may be used in a manual tracking mechanism (Vol. 20) to correct the wind value until the index remains on the echo. It is the tracking operation, rather than the taking of fixes as such, that determines the accuracy required in the circuits producing the index.

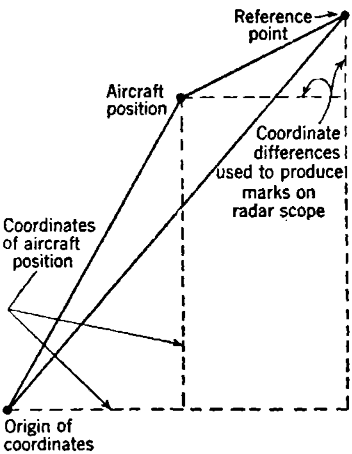


FIG. 7·1.—Coordinates represented in navigational computer.

Alternatives for Producing an Index.—It is first necessary to consider the alternative methods for producing an index on a PPI radar display. One method, used in a GPI produced by the British, is to have a fixed marker at the center of the cathode-ray tube and to move the radar picture in accordance with the input information until a desired echo falls under the fixed index. This has the advantage that rectangular coordinate input information can be used directly; it has the disadvantage that the PPI must be made to give a satisfactorily uniform map of the ground beneath the aircraft. Normally the map will not be uniform, for range from the center of the PPI is ordinarily slant range rather than ground range. This may be corrected by the use of "ground-range sweeps" which compensate for the nonlinearity of the time delay vs. ground-range function by introducing a nonlinear function of radial displacement vs. delay depending on altitude. Even with this sort of correction, however, the method is still relatively inaccurate.

Another method that depends on the cathode-ray-tube display of a true map is the use of a mechanical index. Such an index might be moved in rectangular or polar coordinates. This, in addition to involving the errors of the cathode-ray tube, may involve parallax.

The method chosen in this case is to intensify the beam of the cathode-ray tube in accordance with the GPI information in the same way that the radar information intensifies the beam. Thus the characteristics of the CRT enter in the same way for both radar and GPI information,

FIG. 7·2.—Radar plan-position indicator with range and azimuth markers.

and errors from this source tend to cancel. The procedure will then be to define a time interval measured from the transmitted pulse, which represents GPI slant range; at the end of this interval an intensifying range mark will be produced. To indicate the azimuth of the point given by the GPI coordinates, a radial trace on the PPI will be produced by intensifying the beam for the duration of one or more radial sweeps at the proper antenna angle, that is, when the antenna azimuth relative to north is equal to the angular coordinate corresponding to the GPI information. The index will thus be the intersection of a circle and a radial line.

Circuits Producing the Index.—The computation then consists principally of converting the rectangular coordinates given by the N-S and E-W shaft rotations to the polar coordinates of the radar display. There are several changes of representation associated with this; they are necessitated mainly by the facts that the inputs are shaft rotations, the outputs are to be markers, and the most convenient rectangular-to-polar coordinate transformation uses a-c voltages.

It is necessary to take into account the altitude of the aircraft in producing the range marker. The method used for production of a range marker is intensification of the beam in the cathode-ray tube at a time instant delayed from the transmitted pulse by an interval corresponding to the range of the echo in question. The "delay" of the radar echo is proportional to the "slant range," whereas the GPI shaft rotations measure the components of "ground range."[1] Thus in order to control the slant-range time-modulation circuit it is necessary to correct the ground-range voltage by an amount depending on the altitude.

Equations.—There are only three operations other than the identity operation to be performed in this computation. The first is the rectangular-to-polar coordinate transformation, which may be expressed thus:

$$\left.\begin{aligned} r^2 &= x^2 + y^2, \\ \theta &= \tan^{-1}\left(\frac{y}{x}\right). \end{aligned}\right\} \tag{1}$$

An equivalent set of equations which more closely represents the operation performed is

$$\left.\begin{aligned} x \cos\theta - y \sin\theta &= 0, \\ x \sin\theta + y \cos\theta &= r. \end{aligned}\right\} \tag{2}$$

The second operation to be performed is the computation of the hypotenuse of a right triangle, given the legs.

$$r^2 + h^2 = s^2, \tag{3}$$

where r = ground range,
h = altitude,
s = slant range.

The third operation is the subtraction of shaft rotations necessary to produce the azimuth mark. This mark is to occur when

$$\theta_R = \theta_A + \theta_C, \tag{4}$$

where θ_R = the angle of the resolver shaft (output of rectangular-to-polar transformation),
θ_A = the angle of the antenna with respect to the aircraft,
θ_C = the compass angle (the angle of the aircraft with respect to north).

[1] See Sec. 6·2 and Vol. 20, Chap. 4.

Design Limitations.—In connection with this computer three basic decisions were made:

1. An attempt is to be made to meet strictly the performance specifications and, where applicable, the component specifications of the Army Air Forces and the Navy Bureau of Aeronautics. These specifications relate mainly to the life of the equipment and its performance when exposed to extremes of temperature, humidity, vibration, etc., and are briefly discussed in Chap. 19.
2. An attempt is to be made to employ radical new techniques of electronic construction and the most advanced design techniques available if necessary in order to reduce drastically the weight of this computer.
3. Strong emphasis is to be placed on reliability and the various factors that contribute to it, such as ease of maintenance.

The emphasis on light weight points to the necessity of replacing some components with much lighter substitutes. In some cases components must be redesigned. Several basic electronic devices of this sort which are extremely important in the further development of this computer are lightweight accurate resolvers; small lightweight servo motors, controllable with small tubes; subminiature tubes; lightweight small condensers; and small precision computer transformers. Laboratory development and discussions with manufacturers are initiated or continued at this stage in the computer design, most of these devices having been already under consideration for other purposes or as part of a basic development program.

Errors.—Throughout the design process, the designer must have an idea at least of the order of magnitude of the errors to be expected in computation. In the GPI navigational computer, it is necessary to consider the desired over-all performance in the light of other navigational techniques and of limiting component accuracies. From a consideration of this sort, approximate figures for the permissible over-all error may be specified and this error may be apportioned among the various computing elements.

The discussion of the circuits to be treated here will assume that such a general error analysis has been performed, and that the probable error of a setting of the index, due to the circuits discussed, is not to exceed $\frac{1}{4}$ mile on a radar display having a 50-mile maximum range. An assumption of this sort is quite arbitrary, and, in fact, it is often found that the original apportionment of permissible errors has to be modified as the design proceeds.

7·3. Creating a Block Diagram. *Procedure of Sec.* 2·1.—After the problem has been defined, the general nature of the computer decided upon, and most of the design limitations stated, the next step is to create a block diagram. This process requires ingenuity and some knowledge of available techniques. It is difficult to say when the best arrangement of components for a given problem has been reached; successive trials, aided by experimental and theoretical information, will generally lead to improvements over the first diagrams drawn. Several different representations of quantities will usually be found in the optimum arrangement.

Representation of Quantities.—The nature of the output determines part of the instrumentation; for example, the use of a time-modulated range

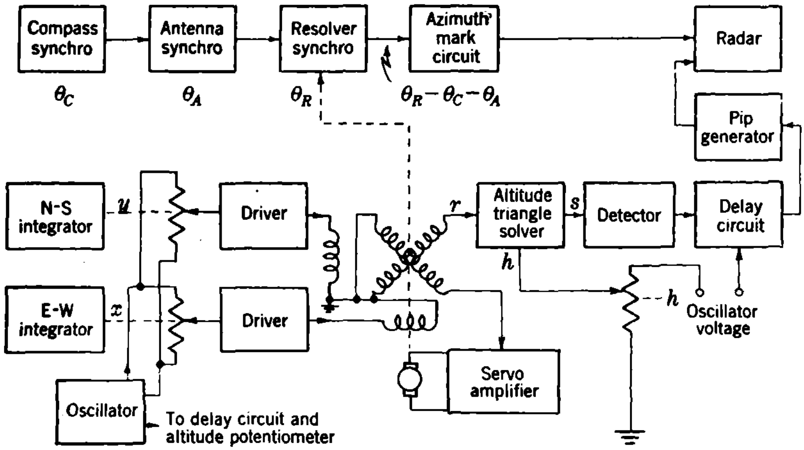


FIG. 7·3.—Generation of range and azimuth markers from integrator outputs.

mark makes it convenient to use a delay circuit (Sec. 6·6) actuated by a d-c voltage. Some sort of transformation from rectangular to polar coordinates is necessary, and at some point, either before or after the transformation, the mechanical input information must be converted to d-c voltage. The principal choice remaining has to do with whether the transformation is to be done mechanically or electrically. In the development of this computer consideration was given to both mechanical and electromechanical methods. An electrical resolver, the electromechanical device, was finally chosen for the operation after it was shown to provide satisfactory accuracy, because it seemed easier to obtain in quantity and because it was lighter than the mechanical device considered.

The block diagram then assumes the form shown in Fig. 7·3. The x- and y-shaft rotations are first converted to alternating voltages by means of linear potentiometers. The voltage across the potentiometer is supplied by a sine-wave oscillator (see below). Since the voltages at the potentiometer arms come from relatively high-impedance sources, precise impedance-changing circuits (drivers) are necessary to reproduce these voltages across the resolver stator windings without loading the potentiometers. The coordinate transformation is done by means of a servo (Sec. 6·9) that turns the resolver rotor until one winding picks up no voltage. At this equilibrium position, the other rotor winding picks up the resultant a-c voltage that measures the ground range r, and the rotor shaft rotation measures the angle $\theta_R = \tan^{-1}(y/x)$.

The range voltage then goes to a triangle solver, which computes to slant range s from r and the altitude. The phase-shift triangle solver,[1] which uses alternating voltages, is particularly convenient here. At this point a rectifier converts s to d-c representation, and the delay circuit produces a proportional time delay. Finally a sharp intensity marker is generated at the time corresponding to s, and this marker is fed to the CRT.

The angle θ_R is taken off by an Autosyn,[2] and by means of differential Autosyns the algebraic sum $\theta_R - \theta_C - \theta_A$ is formed. The rotor output voltage of the final Autosyn is zero whenever $\theta_R - \theta_C - \theta_A = 0$, and at this time the azimuth mark appears. The azimuth mark circuit produces an intensifying pulse when the Autosyn output is a minimum.

An additional amplifier, not shown, is inserted at the output of the resolver. The function of this amplifier is to multiply the r voltage by a constant. This is done in order to permit the resolver drivers and the oscillator supplying voltage to the potentiometers to operate at low voltage level and thereby conserve power. In the entire computer there are several other amplifiers similarly employed so that there is considerable saving of power and consequent reduction of weight in the power supply.

The use of the oscillator voltage as a reference for the linear delay circuit may be done in such a way that, to a first approximation, variations in the oscillator output do not affect the delay of the range mark. If the slope of the triangular waveform is made proportional to the oscillator output, a given range (defined by input shaft rotations) will correspond to a constant delay, regardless of small variations in oscillator voltage.

7·4. Preliminary Design. *Procedure of Sec.* 2·1.—After the block diagram has been drawn, the first stage of design is to select the circuit types to be used. This means taking the design to the point where the values of most of the components are known approximately; the com-

[1] *Cf.* Sec. 6·7.
[2] See Sec. 12·1.

ponents determining the accuracy must be known well enough to permit a performance analysis, and it must be possible either to buy or to manufacture them. When the circuit types to be used are fairly standard, the approximate accuracy obtainable may often be stated before any design work is done for the particular problem at hand. If a new computing circuit is required, however, considerable development work may be required to make sure that the circuit can satisfy the requirements.

Circuit Types.—For each block in the block diagram of Fig. 7·3 a specific circuit type must be chosen. The principal requirement on the *oscillator* that supplies the computing voltages is that it produce a nearly

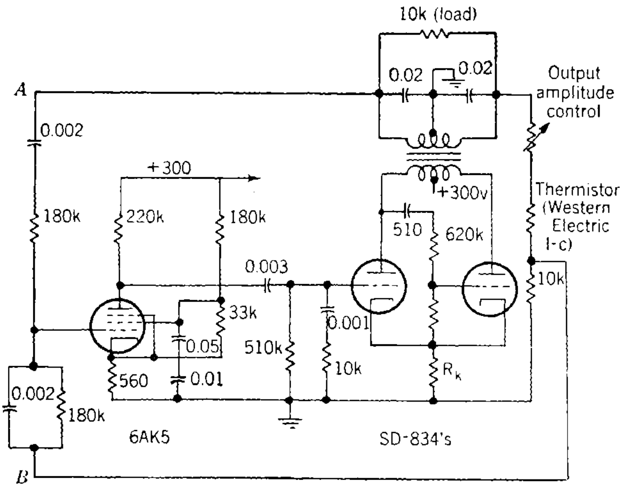


FIG. 7·4.—Oscillator circuit.

pure sine wave, with not more than about 0.1 per cent harmonic content. A further requirement imposed by other circuits in the computer is that the output be fairly constant. Thus the circuit type chosen is a Wien bridge oscillator with a thermistor for amplitude stabilization. Positive and negative feedback are applied at the two terminals *A* and *B* of the Wien bridge. The signal at the input grid is a measure of the unbalance of the bridge. At the fundamental frequency the bridge is balanced, but for harmonics there is considerable unbalance. The result is a sharp frequency characteristic tending to reduce the harmonic content of the output considerably. The feedback is taken from the transformer secondary in order to minimize distortion introduced by the transformer itself.

The *drivers* for the resolver stators are two-stage amplifiers with high feedback gain, accomplishing the same purpose as cathode followers but having less variation in gain with respect to tube change and aging. Circuits of this type are discussed in Vol. 18. The peak variation with respect to tube change is expected to be about ± 0.1 per cent.

The *servo* for the resolver uses an a-c amplifier, phase detector, and final differential-current output stage driving the field windings of a split-

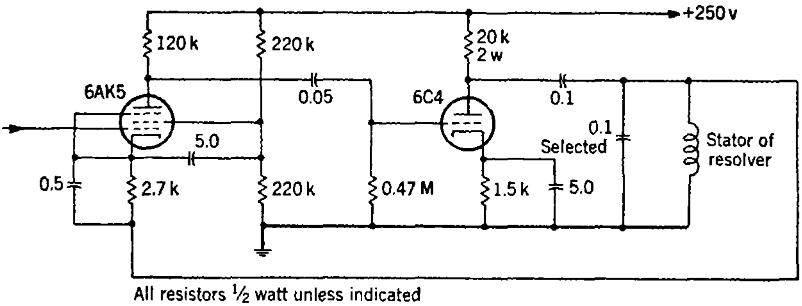


FIG. 7·5.—Resolver driver circuit.

field motor. This circuit may require further development work in order to achieve the error figure of $r\Delta\theta = 100$ yd peak error which was assigned to it for this purpose.

Following the resolver is a *step-up amplifier* which changes the voltage scale by a factor of about 5. This is a three-stage amplifier with a feed-

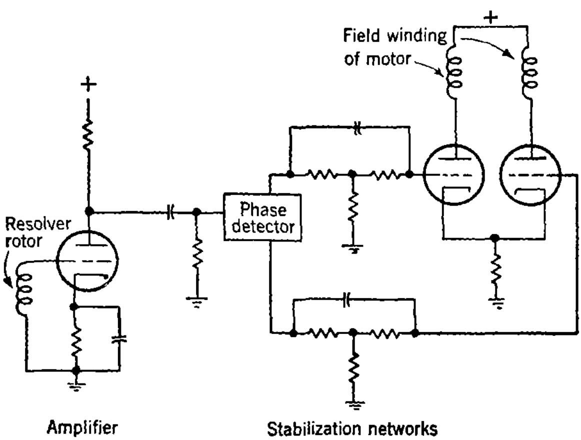


FIG. 7·6.—Resolver servo circuit.

back circuit that consists of a parallel-adding combination of two resistors in the grid circuit. The over-all gain of the amplifier with feedback is determined almost entirely by the values of these resistors. A transformer might be used for this purpose, but the output impedance would be greater, harmonic distortion might be introduced, and there would probably be more variation of ratio with temperature.

The *triangle solver* is of the phase-shift variety discussed in Sec. 6·6, using electronic differentiation and integration to produce a 90° phase shift with amplitude reasonably constant over the frequency range used.

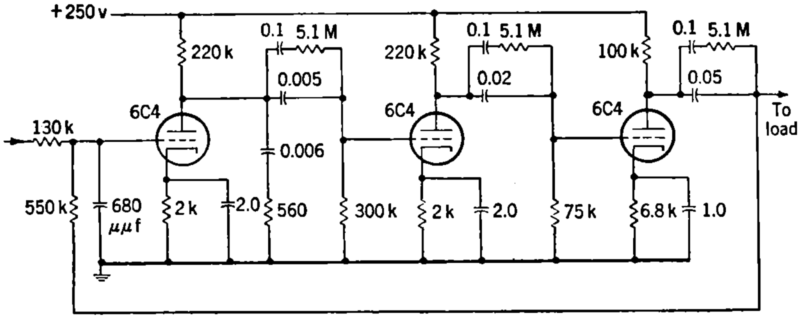


FIG. 7·7.—Step-up driver circuit.

The *rectifiers* (one for each of the two output voltages of the triangle solver) are of the voltage-doubler type, in order to increase the scale of d-c volts per mile at the delay circuit. Crystal rectifiers might be used here to avoid tube drift and to reduce filament power.

After the d-c voltages have been combined in an averaging network, the average is compared with a periodic triangular waveform by means of an amplitude comparison circuit. The triangular waveform is generated by a bootstrap integrator circuit of a type discussed in Vol. 20. Deviations from linearity may be of the order of 0.1 per cent and level shifts 0.2 per cent of maximum output. Methods of producing range marks are also discussed in Vol. 20. A blocking oscillator is used to produce the final sharp pulse.

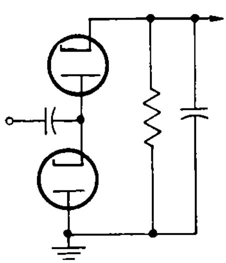


FIG. 7·8.—Voltage-doubler rectifier.

The *azimuth mark circuit*[1] is an amplifier that produces a pulse when the output of the Autosyn chain reaches zero. This output goes to zero twice during every revolution of the antenna, so that some means must be provided for selecting one of these "nulls" and of preventing a mark from being produced by the other, which occurs 180° away from it. The principal sources of error in azimuth marking arise from the Autosyns; there may be errors of approximately 0.2° in the rotor position at which minimum output occurs; and the minimum output may

[1] See Vol. 20, Chap. 4.

differ from zero, so that it is necessary to provide in the amplifier for starting the output pulse before the input reaches zero.

Components.—The *potentiometers* that convert shaft rotation to voltage must have a linear variation of resistance with rotation. For this purpose a 10-turn helical potentiometer[1] can be used. Deviations from linearity are held to 0.1 per cent or less of total resistance. Another possibility is the single-turn RL270 potentiometer (Vol. 17, Chap. 8), which was under development at the time of design and is capable of the same accuracy.

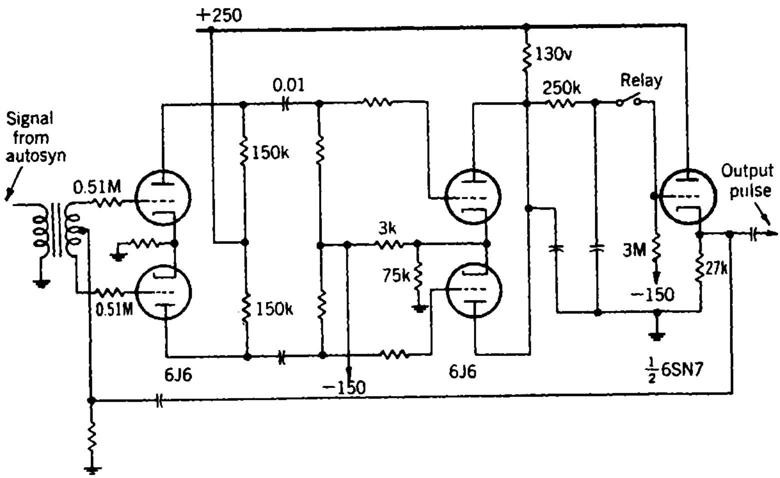


FIG. 7·9.—Azimuth mark circuit.

A *resolver* that meets the design requirements is one developed by Bendix Pioneer (XD-759542) for use in lightweight computers. This device is accurate to 5 min in angle and ± 0.06 per cent of maximum output range under the conditions in which it is used in this computer. The ratio of input to output voltage varies with temperature, changing by about 0.2 per cent over a range of $-55°$ to $+70°$C. This variation may be reduced by changing the electrical loading on the resolver output.

The *precision resistors* that determine the gain of the step-up driver must "track" (preserve the constancy of resistance ratio) over the desired temperature range. These may be wire-wound or other types of precision resistors. The initial adjustment of ratio is made by a factory-set calibration rheostat in series with one of the resistors; thus even if an accuracy of computation of 0.1 per cent is desired, the initial

[1] The "Helipot" made by National Technical Laboratories or the "Micropot" of Thomas B. Gibbs Co.; See Vol. 17, Chap. 8.

values of the resistors need not be held to closer tolerances than 1 per cent.

The constancy of the slope of the triangular waveform with respect to temperature depends on the constancy of the RC product of the integrating network (Vol. 19, Chap. 10). These components must be chosen so that their temperature coefficients cancel to some extent. The accuracy to which this can be done is limited by the designer's knowledge of the temperature coefficients; if standard mica condensers such as those specified in Specification JAN-C-6 (Apr. 20, 1944) are used, the variation of temperature coefficient within one class in the specification is a limitation. The best tolerance specified (type G) is from 0 to -50 ppm/°C. When by proper choice of resistance type the average temperature coefficient of RC is made zero, a change of ± 0.25 per cent over a 100° range can still be expected from a randomly chosen sample. If elements are separately tested and matched, greater accuracy is possible, but quantity production becomes much more difficult.

These, then, are the principal precision components which contribute directly to the error of computation. The tubes used also contribute. Clamp tubes in the linear delay circuit, the diodes in the rectifier, and the coincidence tube all introduce d-c level changes. The triode amplifiers in the oscillator introduce harmonics that must be removed by the Wien bridge feedback.

Other special components introduced in this design are the split-field servo motor, used in the resolver servo; the oscillator output transformer, specially designed for light weight and balanced output; and the thermistor used to stabilize the oscillator output.

If production in quantity is contemplated, the possibility of obtaining or manufacturing all these special components is investigated at an early stage in the design.

Scale Factors.—The maximum range of the radar with which this computer is to be used determines some of the scale factors. The range attainable on ordinary ground echoes (cities, land-water boundaries, etc.) is from 30 to 60 miles; however, in order to provide for the longer ranges that may be obtained if radar beacons are used in navigation, it is preferable to use a maximum range of 100 nautical miles. The potentiometer that converts shaft rotation to voltage must then represent 100 nautical miles[1] in either direction; that is, its entire length corresponds to 200 miles. If single-turn potentiometers with a full rotation of 350° are used, the mechanical input must be at a scale of 200 miles = 350°. For the same reason, the range mark must go out to 100 miles, the delay circuit must be accurate to at least that range, and the full voltage of

[1] In the discussion that follows, "miles" will be used to refer to nautical miles. A nautical mile is equal to a minute of latitude.

the triangular wave form must then correspond to 100 miles. This full voltage is made as large as possible in order that tube drifts shall correspond to small errors in delay.

The plate supply voltage for the amplifiers can be an unregulated 300-volt supply. This limits the possible output of the step-up amplifier following the resolver, for the peak-to-peak swing of the plate of the output tube cannot exceed 300 volts. As a matter of fact, the nature of the circuit restricts the maximum output to about 200 volts peak-to-peak, or 70 volts rms. Thus the a-c voltage scale at the triangle solver is 70 volts rms = 100 miles, or 0.7 volt rms per mile. The voltage scale for circuits preceding the step-up driver is 0.2 (volt rms)/(mile), so that the gain of the step-up driver is about 3.5. The output voltage of the oscillator is then (200 miles)(0.2 volt/mile) = 40 volts rms. This appears across the potentiometers. The maximum voltage in either coordinate that goes to the resolver stator is 20 volts rms.

Since the differentiator and integrator in the triangle solver operate better when the gain is less than unity, a larger altitude voltage than the desired output is fed in. It is desired to provide for altitudes up to 40,000 ft (about 7 miles). If the entire oscillator voltage (20 volts) is used to supply an altitude potentiometer whose full rotation represents 7 miles, a loss in gain of a factor of 14 can be taken in the phase shifter.

In the azimuth data a scale factor might conceivably be used if "two-speed" data from geared-up synchros were used. In this case however, no such gearing is employed, so that in one sense the scale factor for azimuth information may be considered to be unity. The information is actually transmitted in the form of two voltages which measure the projections of a rotating line segment on oblique axes at 120° to one another. One such voltage is used as the input to the azimuth mark circuit. This voltage varies as the sine of output angle, hence in the useful region, where the voltage is nearly zero, it is nearly proportional to the angle itself. A scale of voltage/angle at null may then be defined; if it is measured in volts per radian it is equal simply to the maximum output of the autosyn chain. As the output signal is amplified, this scale factor is multiplied by the gain of the amplifier.

One scale factor that enters into the design, although it does not appear directly in the computation, is the sensitivity of the error signal winding of the resolver. The voltage per radian of error, for small error angles, is equal to the r-voltage that appears across the other rotor winding, assuming that the two rotor windings have equal numbers of turns.

7·5. Performance Analysis. *Procedure of Sec.* 2·1.—The performance analysis is a check on paper to determine in the light of available data whether or not the computer will operate satisfactorily. This consists chiefly of an error analysis with consideration of the effects on errors of the conditions under which the computer is to operate: temperature, humidity, pressure, etc. The particular value of a systematic check of this sort is that it may show that some important data have not yet been taken; for example, the probable life or behavior with temperature of some important component may not be known.

Errors.—In each case when a component error is known, its maximum value is specified. Usually the error varies in such a complicated way that it is extremely inconvenient to give a measure of error such as rms or probable error. A linear potentiometer is a good example of this. Yet when the errors of a number of components are combined it will be extremely rare that the maximum errors of all the components appear simultaneously and all affect the output in the same direction. If the designer is more interested in the probable error of the computer than in the limits of error, he must make some approximations and simplifying assumptions in order to use the data available for the components. At a later stage in the design (after a model has been made) probable errors may be found by taking more data. At the present stage an estimate must be arrived at without this information.

A set of working assumptions that have proved useful are the following:

1. The probable error of a component will be assumed equal to one-third the peak error or tolerance (this corresponds to an error not exceeded in 96 per cent of the cases in a normal probability distribution).
2. Errors may be combined by squaring, adding, and extracting the square root, as is customary for probable errors. The result of calculations carried out on the basis of these assumptions may be expected to be correct within something like a factor of 2. In practice the most serious departures from the behavior predicted in this way have been found to be due to systematic deviations of the central value from the true value, due, for example, to miscalibration.

The approximate values of maximum errors for the various components and circuits mentioned, exclusive of temperature effects, may be tabulated as shown in Table 7·1. These figures may be converted into miles (or degrees in the case of the azimuth error, which may be calculated separately), and the square root of the sum of the squares calculated. If the resulting figure is then divided by 3 (this being the equivalent of dividing each peak error by three), the result will be the assumed probable error.

The range error is then approximately

$$(\Delta r)_{\text{prob}} \approx \frac{\sqrt{(0.2)^2 + (0.1)^2 + (0.06)^2 + 4(0.1)^2 + (.01)^2}}{3} = 0.11 \quad \text{mile.} \tag{5}$$

Similarly, the azimuth error is approximately

$$(\Delta\theta)_{\text{prob}} \approx \frac{\sqrt{(0.08)^2 + (0.06)^2 + 4(0.2)^2}}{3} = 0.14^\circ \tag{6}$$

If a fix is taken at an average range of 50 miles, the corresponding error $r\Delta\theta$ is

$$(r\Delta\theta)_{\text{prob}} = 0.12 \quad \text{mile.} \tag{7}$$

If these range and azimuth errors are combined by rms addition, the result is a probable fix error of 0.18 mile. This is somewhat better than the figure of ¼ mile mentioned earlier.

TABLE 7·1.—MAXIMUM ERRORS

Component	Maximum error	Δr, miles	$\Delta\theta$, degrees
Potentiometers	±0.1% of full range (200 miles)	0.20	......
Drivers	±0.1% of output	0.10	......
Resolver	±5 min; ±.06% of max. output (100 miles)	0.06	0.08
Resolver servo	$R\,\Delta\theta$ = 100 yd at R = 50 miles (average range)	....	0.06
Autosyns	±0.2° each, combining randomly	....	$0.2\sqrt{3}$
Azimuth mark circuit	±0.2°	....	0.20
Step-up driver	±0.1% of output	0.10	
Triangle solver	±0.2% of altitude	0.01	......
Detectors	±0.2 volt (200 volts = 100 miles)	0.10	......
Delay circuit	±0.2 volt (200 volts = 100 miles)	0.10	......
	±0.1% of full range (100 miles) departure from linearity.	0.10	......

Other Considerations.—The performance analysis of this computer with respect to temperature is not detailed here. Some important sources of error may be mentioned however. The variation with temperature of the stator-to-rotor voltage ratio of the resolver (0.2 per cent) is significant and is characteristic of transformers. Change of air gaps with expansion of core metal and change of wire resistance with temperature are probably responsible for the effect. These variations may be compensated once they are accurately known.

At this stage of the design it is also well to consider the interaction of the various separate circuits. For example, the loading effect of the resolver drivers on the precision potentiometers may be checked. The input impedance of the drivers is very high and chiefly capacitive; but until a rough quantitative check has been made, the designer cannot be sure that significant phase shifts will not occur. The interaction of the x and y channels may be examined in case current in a rotor winding should produce coupling between the two stators. The operation of the step-up driver and triangle solver with the detectors should be checked, for the pulses of current drawn by the detectors may produce undesired transients in the preceding circuits. These points require experimental work with two circuits at a time and may expose difficulties before a model has been constructed.

7·6. Detailed Design. *Procedure of Sec.* 2·1.—The design must now be carried to the point where each component is well enough specified so that it may be ordered and so that it can be counted on to function properly in the computer. In the following description, circuit designs will be given, but in many cases tolerances are not available.

The tubes shown in these designs are miniature tubes (the 6C4 triode, 6AK5 pentode, and 6AL5 double diode) and subminiature tubes (the Sylvania SD-834 triode[1] and Raytheon CK-604 pentode). The circuits using miniature tubes may be redesigned for other tube types by means of relatively minor changes.

Oscillator.—The 500-cps oscillator, a Wien bridge circuit, is based on a design made at Bell Telephone Laboratories with the assistance of a Radiation Laboratory engineer. The circuit is shown in Fig. 7·4. The output is 40 volts rms across 10,000 ohms. Tests on the circuit indicated only 0.05 per cent second harmonic and 0.1 per cent third harmonic in the output waveform.

Resolver Drivers.—These drivers are two-stage amplifiers with cathode feedback, as shown in Fig. 7·5. The design of these circuits is discussed in detail in Vol. 19. With respect to variation in tubes, load impedance, condensers, and resistors, the variation of gain does not exceed ±0.1 per cent.

Resolver Servo.—For the servoamplifier to be used with the split-field motor, an a-c amplifier, phase detector, "phase-lead" network, and a differential current output stage, as shown in Fig. 7·6, should prove satisfactory but is untested. A similar circuit which was used in a related equipment is described in Sec. 14·3.

Step-up Driver.—The high feedback gain of this circuit and the use of three stages of amplification necessitate careful design to prevent oscillation. The design procedure for this amplifier is discussed in detail in Vol. 18. The final design is shown in Fig. 7·7. The use of a

[1] Now the 6K4.

step-up amplifier here allows lower voltages to be used in the preceding stages, with resultant over-all savings in power and weight.

Triangle Solver.—A tentative circuit design, incorporating the reduction in gain previously mentioned, is discussed in Sec. 6·6. The sources of error in this type of triangle solver are also discussed in that section.

Detectors.—The type of voltage doubler recommended is shown in Fig. 7·8. The principal design problem with this type of detector is usually selection of proper capacitor values; this should not be difficult in the present case, because of the low output impedance of the step-up driver.

Time-modulation Circuit.—A time-modulation, or "delay," circuit, including sweep generator, coincidence circuit, and pip generator, is similar to those for which detailed discussion and design procedures are given in Vol. 20.

Azimuth Mark Circuit.—This circuit is shown in Fig. 7·9. It includes a differential amplifier, a plate-circuit detector, and a regenerative loop which causes the output to be a rectangular gate several milliseconds in length. Differential rather than single-ended amplification is used in order to make it possible to use a shorter time constant in the detector. This is necessary because the angular velocity of the antenna (200°/sec) is such that one cycle of the 400-cps line supplying the autosyns corresponds to $\frac{1}{2}°$.

The final autosyn is connected in such a way as to make available a voltage that goes (approximately) to zero only *once* each revolution, rather than twice. This is done by adding to the sine-modulated output voltage a constant a-c voltage equal to the maximum output. The resulting voltage is used to remove the "back trace" of the azimuth mark.

7·7. Finishing the Design.—Several steps remain to be done in the design of this computer. The detailed design has to be carried out for several of the circuits. A model must be built as a check on the combination of circuits and mechanical components. One difficulty often encountered when circuits and mechanical parts are to be designed under pressure of time is that the mechanical design has to be "frozen" much sooner than the electrical design. This means that last-minute changes can be made in the electrical design if experiments show that some expectations are not realized; the corresponding mechanical changes, however, may be much more difficult to make.

It is in the remainder of the design that the limitations already mentioned—light weight, reliability, conformity to aircraft specifications—enter particularly. Care must be taken in chassis layout to save weight and space but at the same time to make servicing possible and to avoid excessive heat dissipation at "hot spots" in the chassis. Exhaustive tests must be made to see whether or not the circuits perform satisfactorily with respect to temperature, vibration, humidity, etc. Even such things as the effect of chassis warm-up should be considered, for it has been found in some instances that a circuit calibrated a short time after it has been turned on may show substantial systematic errors after an hour or two of operation.

SPHERICAL COORDINATE INTEGRATION

7·8. Statement of Problem and Preliminary Design Information.—The spherical coordinate integrator discussed in the following sections was designed as part of a radar training device and solves the differential equations of motion of an aircraft as observed from a moving ship. Information such as heading, air speed, rate of climb, rate of turn of the aircraft, direction and magnitude of the wind, and the course, speed, and rate of turn of the ship are set into the integrator, which then operates upon this information in such a manner as to yield the position of the aircraft with respect to the ship, as measured in spherical coordinates. Thus, slant range, azimuth angle measured in the horizontal plane, and elevation angle of the aircraft above the horizontal plane are computed as continuous functions of time. These data are used to simulate the position information ordinarily obtained by a radar mounted aboard the moving ship and tracking the maneuvering aircraft.[1]

Two fundamental methods of solving the differential equations of motions may be investigated. One method involves resolution of the various velocity vectors into three mutually perpendicular components whose directions are fixed with respect to the earth, integration in these coordinates, and transformation of the resulting position information into the desired spherical coordinates. The second method consists of transformation of the various velocity vectors into vectors in spherical coordinates corresponding to the rate of change of range of the aircraft, rate of change of azimuth, and rate of change of elevation, respectively, followed by integration in these coordinates, no coordinate transformation of the resulting position information being necessary. After careful consideration of both methods, the latter one was chosen. It is simpler than the first method, since no final conversion from one set of coordinates to another is used; because of this it is also capable of giving smoother output information, since the output of a velocity servo is usually smoother than that of a position servo running at a comparable velocity and is certainly smoother than that of a velocity servo and a position servo in series. This smoothness is desirable from the standpoint of the radar trainer application.

[1] A complete description of a trainer employing the spherical coordinate integrator will be found in "SP Trainer," Radiation Laboratory Report No. 928.

Upon investigation, several assumptions are found that simplify the general equations and are acceptable from the radar training standpoint, since they introduce negligible errors. These simplifying assumptions resulted in considerable savings in over-all system complexity. The effects of wind and ocean currents on the ship may be assumed equal to

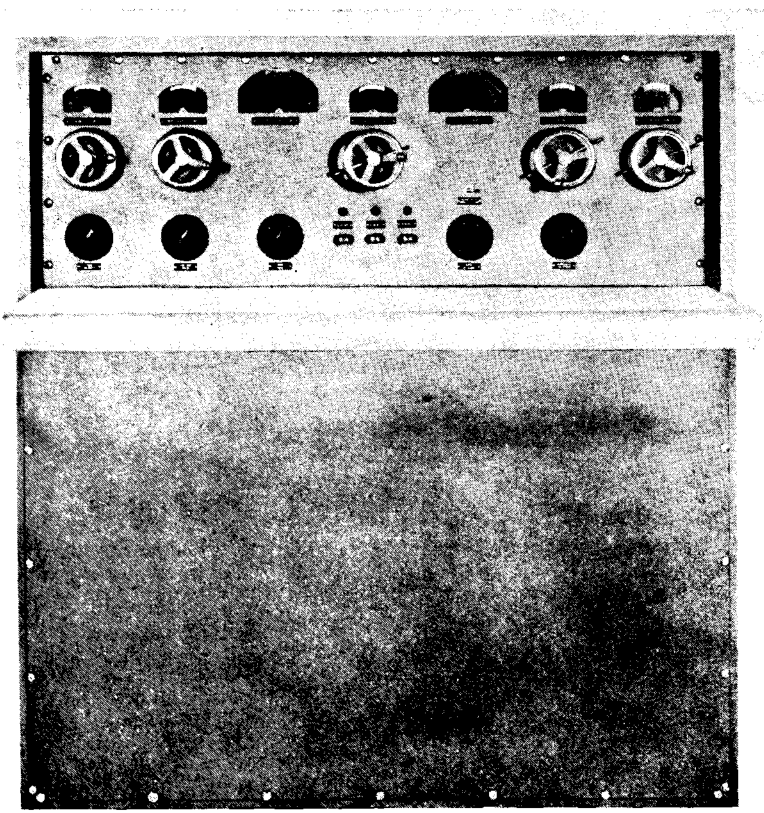


FIG. 7·10.—Radar trainer using spherical coordinate integrator.

zero. Curvature of the earth may be neglected. Skidding of the aircraft and the ship during turns may be neglected. Since a stabilized antenna is used in the radar equipment that this trainer component was to simulate, roll and pitch of the ship may be neglected.

With these assumptions, the detailed differential equations of motion of the aircraft with respect to the ship can now be developed.

The simplified system geometry is shown in Fig. 7·12. Complete

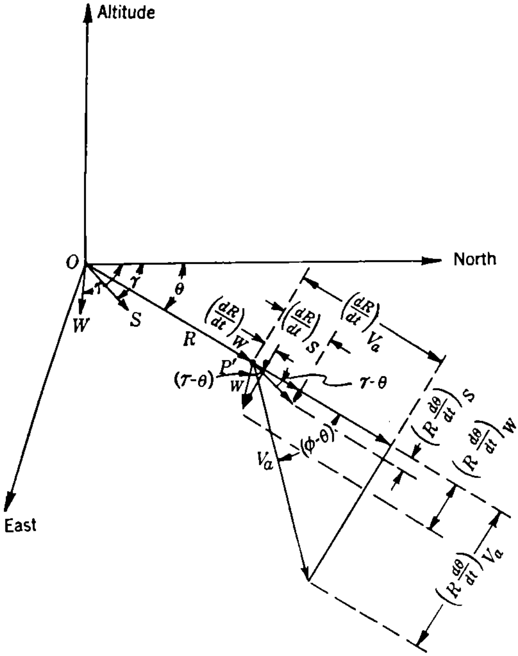


FIG. 7·11.—Ground plane geometry including wind and site motion.

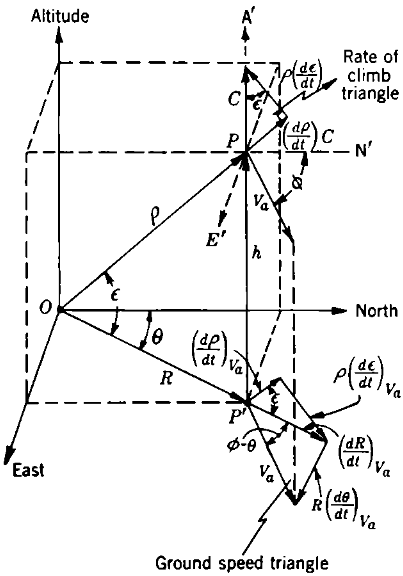


FIG. 7·12.—Simplified system geometry.

ground plane geometry, including the effects of wind and motion of the ship, is shown in Fig. 7·11.

The symbols used throughout the discussion have the following meaning:

ϵ = elevation angle of aircraft with respect to horizontal,
θ = azimuth angle of aircraft with respect to north,
ϕ = heading angle of aircraft with respect to north,
γ = course of ship with respect to north,
τ = direction of wind with respect to north,
C = rate of climb,
V_a = horizontal true air speed,
S = ship speed,
W = wind speed,
h = altitude of aircraft,
ρ = slant range,
R = ground range,
O = position of radar,
P = position of aircraft,
P' = projection of aircraft position on ground plane,
t = time.

Note that V_a is a horizontal speed. When the aircraft is climbing, the true air speed is the vector sum of V_a and C.

Figure 7·12 indicates the position of the aircraft in space relative to the ship at a given time t as well as the velocities of the aircraft at that instant. Since the solution is desired in spherical coordinates, the input information pertaining to the motions of the aircraft, ship, and wind must be resolved into vectors representing rates of change of range, elevation angle, and azimuth angle.

By inspection of Fig. 7·12, the radial component of ground speed is

$$\left.\frac{dR}{dt}\right)_{V_a} = V_a \cos(\phi - \theta), \tag{8}$$

where the subscript V_a indicates the component is due to horizontal air speed. Similar components due to ship motion and wind are obtained as shown in Fig. 7·11.

Since

$$\left.R\frac{d\theta}{dt}\right)_{V_a} = V_a \sin(\phi - \theta), \tag{9}$$

the rate of change of azimuth angle due to the (horizontal) air speed is given by

$$\left.\frac{d\theta}{dt}\right)_{V_a} = \frac{1}{R} V_a \sin(\phi - \theta). \tag{10}$$

Here again similar components are obtained as a result of the ship motion and the presence of wind. The total rate of change of azimuth angle is given by

$$\frac{d\theta}{dt} = \left.\frac{d\theta}{dt}\right)_{V_a} + \left.\frac{d\theta}{dt}\right)_{W} - \left.\frac{d\theta}{dt}\right)_{S} \tag{11}$$

and upon integration, the actual azimuth angle of the aircraft is obtained.

In a similar manner the rates of change of range and elevation angle are obtained. Substitution yields the desired differential equations of motion:

$$\left.\begin{aligned} \frac{d\theta}{dt} &= \frac{1}{R}\,[V_a \sin(\phi - \theta) + W \sin(\tau - \theta) - S_s \sin(\gamma - \theta)] \\ \frac{d\epsilon}{dt} &= \frac{1}{\rho}\,\{C \cos\epsilon - [V_a \cos(\phi - \theta) + W \cos(\tau - \theta) \\ &\qquad - S \cos(\gamma - \theta)] \sin\epsilon\} \\ \frac{d\rho}{dt} &= \{C \sin\epsilon + [V_a \cos(\phi - \theta) + W \cos(\tau - \theta) \\ &\qquad - S \cos(\gamma - \theta)] \cos\epsilon\} \end{aligned}\right\} \tag{12}$$

In order to obtain the actual displacements, these differential equations must be integrated with respect to time to give

$$\left.\begin{aligned} \theta &= \theta_0 + \int_{t_0}^{t} \frac{1}{R}\,[V_a \sin(\phi - \theta) + W \sin(\tau - \theta) \\ &\qquad - S \sin(\gamma - \theta)]\, dt \\ \epsilon &= \epsilon_0 + \int_{t_0}^{t} \frac{1}{\rho}\,\{C \cos\epsilon - [V_a \cos(\phi - \theta) + W \cos(\tau - \theta) \\ &\qquad - S \cos(\gamma - \theta)] \sin\epsilon\}\, dt \\ \rho &= \rho_0 + \int_{t_0}^{t} \{C \sin\epsilon + [V_a \cos(\phi - \theta) + W \cos(\tau - \theta) \\ &\qquad - S \cos(\gamma - \theta)] \cos\epsilon\}\, dt \end{aligned}\right\} \tag{13}$$

These are the general equations[1] which are to be solved by the integrator. The quantities with the 0 subscript indicate the initial displacements at the time t_0 when the input data are introduced.

7·9. Integrator System Operation.—A simplified block diagram of the integrator is shown in Fig. 7·13. The manner in which the necessary input data are entered and the way in which the integrator solves the equations can be explained with the aid of the block diagram. A more detailed description of the operation of the individual blocks is presented in a later section.

The input knobs that are employed to enter the necessary rate data

[1] A detailed development of these equations of motion will be found in the previous reference.

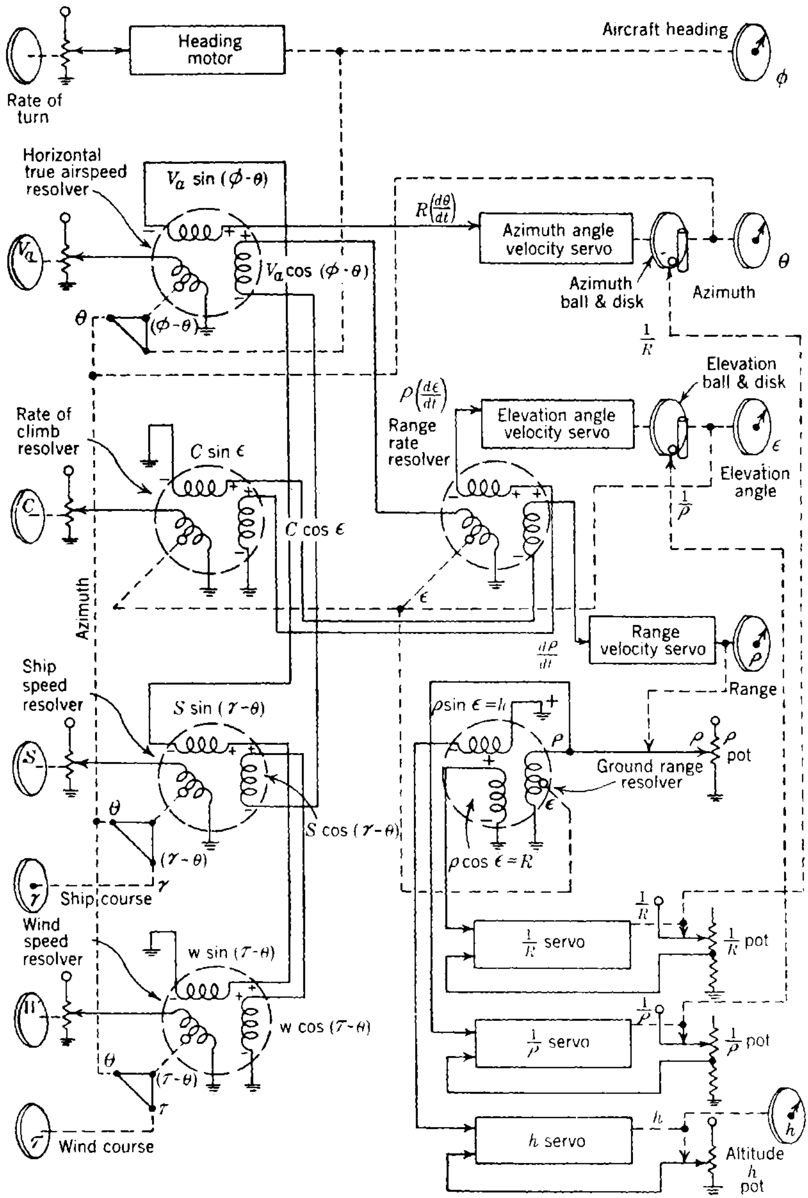


FIG. 7·13.—Simplified block diagram.

are shown at the left of the figure, while the output dials representing the spherical coordinates of the aircraft with respect to the ship are shown to the right.

The input rates are converted from mechanical shaft displacements to a-c voltages whose magnitudes are proportional to the respective rate shaft displacements. The conversion element in each case is a linear potentiometer fed from a constant-voltage a-c source.

The rate-of-turn voltage derived from the rate-of-turn input knob is used to control the speed of an a-c motor. Over the range used, the speed of the motor is roughly linear with control voltage so that the output shaft rotates the compass heading dial and shaft at a speed roughly proportional to the desired rate of turn. The aircraft heading dial in measuring the displacement integrates the rate of turn. This crude velocity control can be used only because the required rate-of-turn accuracy is not great.

The (horizontal) air speed knob is geared to the arm of a linear potentiometer which is electrically connected to the rotor of the air speed resolver. A resolver develops two output voltages proportional respectively to the product of the sine and cosine of the rotor angle and the rotor input voltage. The rotor is turned by a mechanical differential which has the aircraft heading ϕ as one input and the azimuth angle θ as the other. The differential is connected so that the output shaft turns as $(\phi - \theta)$. If the voltage impressed on the rotor of the true air speed resolver is proportional to V_a, the output voltages are proportional to $V_a \cos (\phi - \theta)$ and $V_a \sin (\phi - \theta)$. These voltages are indicated on Fig. 7·13 alongside the corresponding output winding of the resolver. This resolver, therefore, has solved the ground speed triangle of Fig. 7·12.

The rate-of-climb input knob is geared to the arm of a linear potentiometer which is electrically connected to the rotor of the rate of climb resolver. The rotor is geared to the elevation angle shaft; and since the rotor input voltage is proportional to C, the output voltages are proportional to $C \sin \epsilon$ and $C \cos \epsilon$ respectively.

In a similar manner, the ship speed resolver takes the input voltage S and the rotor angle $(\gamma - \theta)$ to give output voltages proportional to $S \cos (\gamma - \theta)$ and $S \sin (\gamma - \theta)$. The wind speed resolver takes the input voltage W and the rotor angle $(\tau - \theta)$ to give output voltages proportional to $W \cos (\tau - \theta)$ and $W \sin (\tau - \theta)$.

By referring to Eq. (12), it is seen that we must take sums and differences of the vector quantities already obtained and further operate on the resultants. This is done by connecting the resolver output windings in series. Addition is performed by connecting the windings in like phase, and subtraction is performed by connecting the windings in phase opposition. The relative phases are shown in Fig. 7·13 by the plus and minus signs at each terminal of the resolver output windings.

By following the series circuits starting with the wind speed resolver output windings and paying proper regard to the phasing, we obtain a voltage proportional to

$$W \cos(\tau - \theta) - S \cos(\gamma - \theta) + V_a \cos(\phi - \theta) = \frac{dR}{dt} \tag{14}$$

which feeds the rotor of the range rate resolver.

The second series circuit yields a voltage proportional to

$$W \sin(\tau - \theta) - S \sin(\gamma - \theta) + V_a \sin(\phi - \theta) = R\frac{d\theta}{dt} \tag{15}$$

which is employed to drive the azimuth angle velocity servo. Equation (14) indicates that we have obtained an a-c voltage proportional to the algebraic sum of all components entering into the rate of change of ground range. Similarly Eq. (15) indicates that we have obtained an a-c voltage proportional to the algebraic sum of all components entering into the rate of change of azimuth angle. It still remains, however, to divide Eq. (15) by R in order to obtain the actual rate of change of azimuth angle $d\theta/dt$.

As shown in Fig. 7·12, the quantity dR/dt must be further resolved to introduce the elevation angle ϵ. This is done by the range rate resolver as shown in Fig. 7·13. The voltage obtained in Eq. (14) feeds the rotor of the resolver, while the angle ϵ is set into the resolver by the shaft. Hence, we obtain voltages proportional to $(dR/dt) \cos \epsilon$ and $(dR/dt) \sin \epsilon$, respectively. The contributions to the motion of the aircraft resulting from the rate of climb are entered by adding the respective components from the rate-of-climb and range rate resolvers as shown in Fig. 7·13. Thus, by following through the two series circuits in a manner similar to that used previously, we have

$$C \cos \epsilon - \frac{dR}{dt} \sin \epsilon = \rho \frac{d\epsilon}{dt} \tag{16}$$

and

$$C \sin \epsilon + \frac{dR}{dt} \cos \epsilon = \frac{d\rho}{dt} \tag{17}$$

Equation (16) must be divided by ρ and integrated in order to give the movement of elevation angle resulting from the input data. Similarly, if Eq. (17) is integrated, the increment in range will have been obtained, and the increment in azimuth can be obtained by dividing Eq. (15) by R and integrating. The above quantities which are to be integrated are all expressed by voltage amplitudes.

Let us follow the integration of Eq. (15). It is desired to obtain

$$\theta = \theta_0 + \int_{t=0}^{t=t} \frac{1}{R} \left[R \frac{d\theta}{dt} \right] dt, \tag{18}$$

where the quantity $R\ (d\theta/dt)$ is now available as a voltage $e(t)$. We therefore write

$$\theta = \theta_0 + \int_0^t \frac{1}{R} e(t)\, dt. \tag{19}$$

Let

$$u = \int e(t)\, dt$$

and

$$\theta = \theta_0 + \int_{t=0}^{t=t} \frac{1}{R}\, du\cdot$$

The integration is actually broken into the same two mathematical steps, as above. The voltage $e(t)$ is first integrated with respect to time by a velocity servo integrator. The function $1/R$ is then integrated with respect to the output of the velocity servo integrator, by means of a ball-disk integrator. The ball-disk integrator used for this second integration is a convenient device for integrating with respect to a variable other than time, whereas the velocity servo is a convenient method of integrating with respect to time. A more complete discussion of this distinction will be found in Chap. 4.

An alternate method of solving the integration problem would have been to multiply the voltage $e(t)$ by $1/R$ before integrating by the velocity servo, the ball-disk integrator not being used. This alternative method was not feasible here because a speed range of 10^6 (ratio of fastest speed to slowest speed) was required. Neither a velocity servo nor a ball-disk integrator is normally capable of such a wide speed range. However, by cascading two devices, each with a speed range of 10^3, the over-all speed range requirement could be met. In an exactly analogous fashion ϵ and ρ are obtained.

In connection with the integration of Eqs. (15) and (16), the multiplying factors $1/R$ and $1/\rho$, respectively, must be used to position the balls of the ball-disk integrators described above. These functions are entered by dividing servos, which receive as their input data voltages proportional to R and ρ, respectively. The operation of these servos is shown in Fig. 7·13. The feedback voltage closing the loop of each servo and used to balance the input voltage is derived from the potentiometer on the servo output shaft, connected as shown in Fig. 7·13. A constant a-c voltage is applied to the arm of this potentiometer, and hence, by Ohm's law, the current through the lower part of this potentiometer and its associated small resistance is inversely proportional to the sum of the small resistance and the included resistance of the potentiometer. The voltage across this small resistor is then balanced against the input voltage in each servo, leading to a null in one case at a resistance proportional to $1/R$ and in the other case to $1/\rho$. The mechanical motion is used directly to adjust the ball position of the ball-disk integrator.

Since the elevation angle ϵ has been obtained and the range ρ is known, the ground-range altitude triangle may be solved by means of

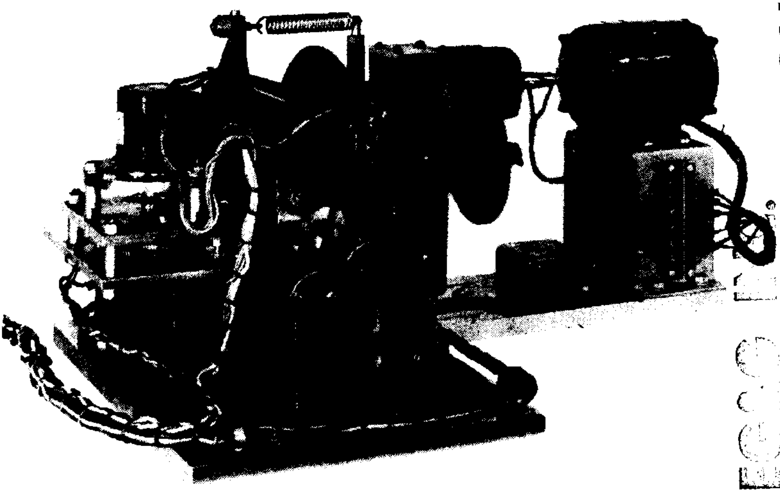


FIG. 7·14.—Combination velocity servo and ball-disk integrator.

the ground-range resolver. If the rotor of this resolver is fed from the arm of the ρ potentiometer and the ϵ shaft is geared to the rotor, the output voltages will be proportional to the ground range and altitude respectively, since

$$R = \rho \cos \epsilon, \tag{20}$$

and

$$h = \rho \sin \epsilon. \tag{21}$$

Thus an a-c voltage whose amplitude is proportional to ground range R is obtained and is used as the reference voltage to obtain the displacement proportional to $1/R$.

It is interesting to note that the ground range R could have been obtained in a different way. It will be recalled that a voltage proportional to dR/dt was obtained in order to feed the rotor of the range rate resolver. Since this quantity represents the rate of change of ground range, upon integration there would be obtained the ground range R. Although this method might seem more straightforward than the one described above, it suffers from the fact that for the same over-all accuracy greater accuracy is required of the individual elements than is the case

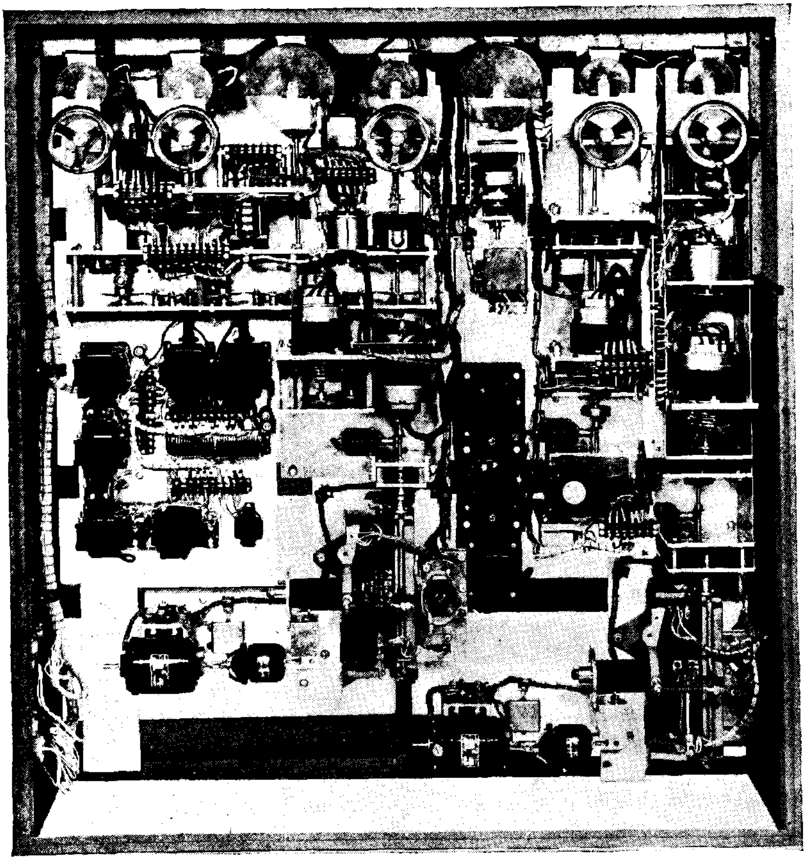


FIG. 7·15.—Radar trainer with front panels removed.

with the method actually used. With the present system if errors arise in the integration of dR/dt and $d\epsilon/dt$ to give ρ and ϵ respectively, the ground range R will still be consistent with these output data, since it is derived from them. This is a typical example in which the choice of a block was made on the basis of the effects that it would produce elsewhere in the system and involves the concept of error cancellation discussed in Sec. 2·8.

Since the ground-range resolver was introduced to obtain a voltage proportional to ground range in accordance with Eq. (20), at the same time there can be obtained a voltage whose amplitude is proportional to the altitude of the aircraft. This is given by Eq. (21).

A linear potentiometer which is driven from a constant a-c voltage source serves as the data output element. The arm of this potentiometer is geared to the servo motor as well as to the altitude-indicating dial and is electrically connected to the input circuit of the servoamplifier. The altitude voltage obtained from the ground-range resolver is also fed into the input circuit of the servoamplifier. In the usual manner, the servo motor will rotate the potentiometer arm in a direction such that the two input voltages are equalized.

The terms ρ_0, θ_0, and ϵ_0, which appear in the integrator equations, are set in manually as initial displacements. These terms are integration constants that fix the position of the aircraft with respect to the moving ship at the start of a trainer problem.

The above analysis of the system operation by means of the simplified block diagram has shown how the range, elevation angle, azimuth, and altitude of a moving aircraft with respect to a moving ship can be obtained. The section that follows will present the actual detailed circuits used and some of the design problems encountered.

7·10. Unit Operation.—This section will discuss specific circuit details and design problems encountered during the development of the spherical coordinate integrator. The values of circuit elements actually used are given in the figures so that the reader can obtain an idea of how the theory presented in the previous section was reduced to practice.

Figure 7·16 is a schematic diagram of the rate-of-turn channel. A Diehl FPE-25 two-phase induction motor is used as the turn rate motor. The rate-of-turn potentiometer, which is fed from a center-tapped autotransformer, controls its speed. One winding of the motor receives constant excitation from the 60-cps line, while the other winding receives variable power from the rate-of-turn potentiometer and the autotransformer center tap. The capacitor placed in series with the fixed or constantly excited field winding shifts the phase of the excitation to provide the quadrature fields necessary for motor operation.

The voltage applied to the control field is a minimum when the arm of the rate of turn potentiometer is at the center and is a maximum at either end. Since the voltage is referred to the center tap of the autotransformer, the phase reverses as the potentiometer arm sweeps from one end through the center to the other. In this manner, a voltage reversible in phase and adjustable in amplitude is applied to the control field of the motor. The direction of rotation of the motor is dependent upon the phase of the control voltage, and its speed is roughly proportional to the voltage amplitude in the range covered.

Since the motor requires a control voltage of the order of 15 volts to start under the actual loading conditions, a shaft rotation of about $\pm 40°$ from the center would ordinarily be required to start the motor. By adjusting the resistance of the zero set potentiometer, the voltage at the center of the rate-of-turn potentiometer can be made to equal the value necessary to just start the motor. Hence, by modifying the standard commercial potentiometer, the rate-of-turn motor can be actuated with about $\pm 1°$ rotation of the potentiometer arm. Although this feature in no way alters the theory or manner of operation, it does provide an improvement in the method of turn simulation.

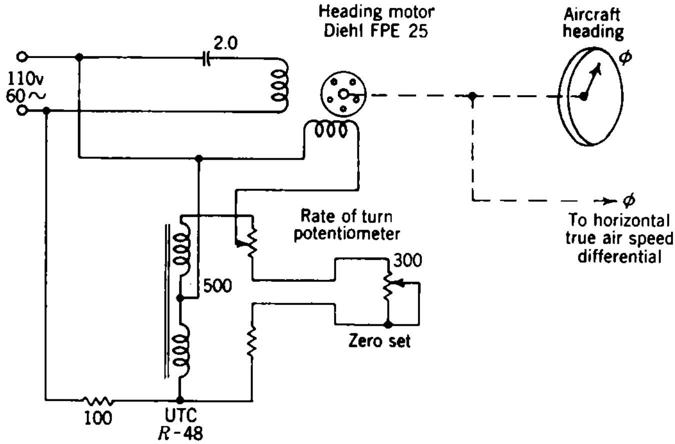


Fig. 7·16.—Rate-of-turn control.

The schematic circuit diagram for the resolver channel is given in Fig. 7·17. The central power source is a center-tapped autotransformer fed from the 60-cps line. The horizontal true air speed knob is geared to the arm of the true air speed potentiometer. The potentiometers P_1 and P_2 are included so that the maximum and minimum true air speeds respectively can be preset. The rotor winding of the true air speed resolver is fed from the secondary of a step-down transformer which in turn is driven by the air speed potentiometer.

The step-down transformer is used so that the loading of the potentiometer by the rotor of the resolver is reduced to a negligible value. The load impedance presented to the potentiometer is essentally that of the reflected rotor impedance, which is high enough to prevent excessive loading. Since the load is inductive and the source impedance is resistive, a phase shift occurs that cannot be tolerated for reasons that

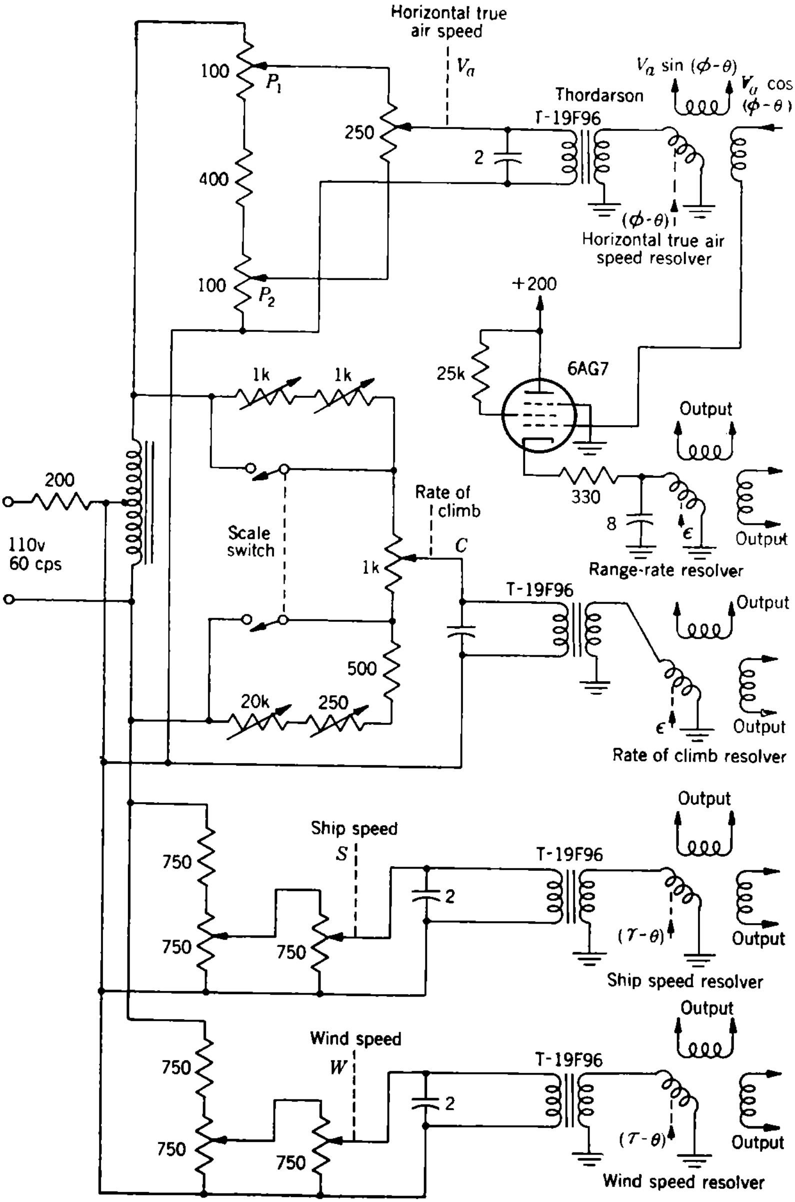


FIG. 7·17.—Resolver channel.

will be discussed later. This phase shift is a function of potentiometer setting because the source resistance varies as the arm rotates. By tuning the load circuit comprising the step-down transformer, resolver, and secondary circuit of the resolver, a load that is higher and resistive can be presented to the true air speed potentiometer. In this manner, approximately zero phase shift can be maintained independent of potentiometer setting. The 2.0-μf capacitor in parallel with the transformer primary is used for this tuning. It has been found that commercial 10 per cent tolerance capacitors maintain the over-all circuit phase shift within the desired limits.

The rate-of-climb resolver is fed in a similar way, although the resistor network required to feed the rate-of-climb potentiometer is quite different. The rate-of-climb voltage must be able to reverse phase, one phase indicating a climb and the other phase indicating a dive. In addition, the maximum dive voltage must be greater than the maximum climb voltage, since an aircraft may dive at much greater speeds than it climbs. The four potentiometers can be adjusted to provide for the proper climb and dive voltages as well as providing for a dual scale type of presentation. With the switch in Position 1, a regular scale is provided which is used for most applications; however, Position 2 may be used to give lower rates of climb or dive for the same potentiometer shaft rotation, providing greater accuracy when setting in low rates of climb or dive. The phase of the control voltage reverses when the arm is in the electrical center of the potentiometer network in a manner similar to that discussed above for the rate-of-turn potentiometer.

A step-down transformer plus tuning capacitor is again employed to minimize loading of the rate-of-climb potentiometer and phase shifts, as discussed above.

The ship and wind speed resolver circuits are made identical, since the maximum magnitude of wind considered is in the order of the maximum ship speed. Potentiometers P_7 and P_8 are included so adjustment of the respective maximum speeds can be made if desired. The resolvers are driven in the same manner as those discussed above so that potentiometer loading and phase shifts are kept to a minimum. The fixed resistors in series with P_7 and P_8 respectively are included to attenuate the voltage to the proper level for use in the two channels. The relative amplitudes of the voltages impressed on the rotors of the resolvers must correspond to the relative amplitudes of the quantities that they represent. These relations are set by the values of resistors and potentiometers used in the networks feeding the rate potentiometers.

In the discussion of the block diagram, it was stated that the range rate resolver received its voltage from one of the output windings of the true air speed resolver. This cannot be done directly in practice, since the resolvers used must be open-circuited in order that the summation of the voltage components be correct. To prevent loading, a 6AG7 cathode follower is used as a buffer and driver for the range rate resolver rotor. The rotor winding is placed directly in the cathode circuit although it is tuned by a 8.0-μf capacitor to obtain a higher and resistive cathode impedance.

With the Diehl resolvers used, a 2/1 step-up exists from rotor to stator windings. Since the same ratio exists in all previously mentioned resolvers, no serious effects are introduced; however, the voltages on the output windings of the range rate resolvers are the result of two cascaded resolvers operating on the input signals. In order to correct for this

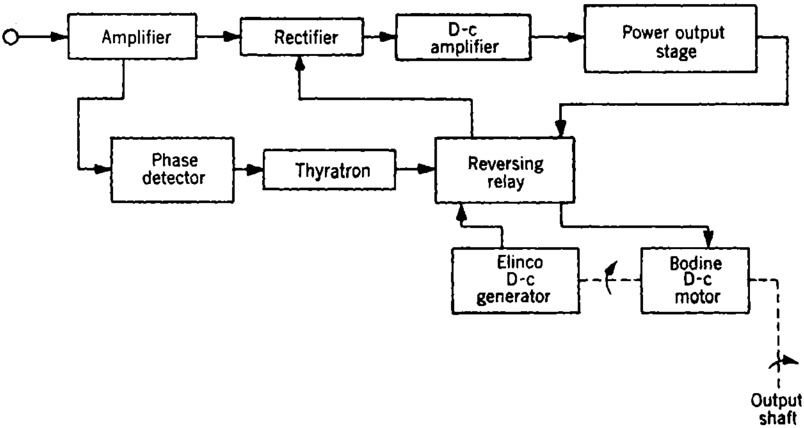


FIG. 7·18.—Velocity servo block diagram.

discrepancy in relative scale factors, the actual voltage impressed on the rotor of the range rate resolver is only one-half the output voltage from the true air speed resolver. The attenuation is accomplished by the 330-ohm resistor placed in series with the rotor winding. Thus, the cathode-follower stage serves both as an impedance matching element and as an attenuator to equalize the existing scale factors.

The interconnections of the respective stator and rotor windings are as shown in Fig. 7·13. The input impedance of the velocity servos is large, since the signal is applied directly to the control grid of a vacuum tube, no grid resistor being necessary.

The series circuits comprising the resolver stator windings generate the correct voltages only if negligible current flows. If current passes through the windings, a voltage drop occurs that causes an error proportional to the winding impedance and current magnitude. In the present case, however, the current flowing is essentially zero, since the

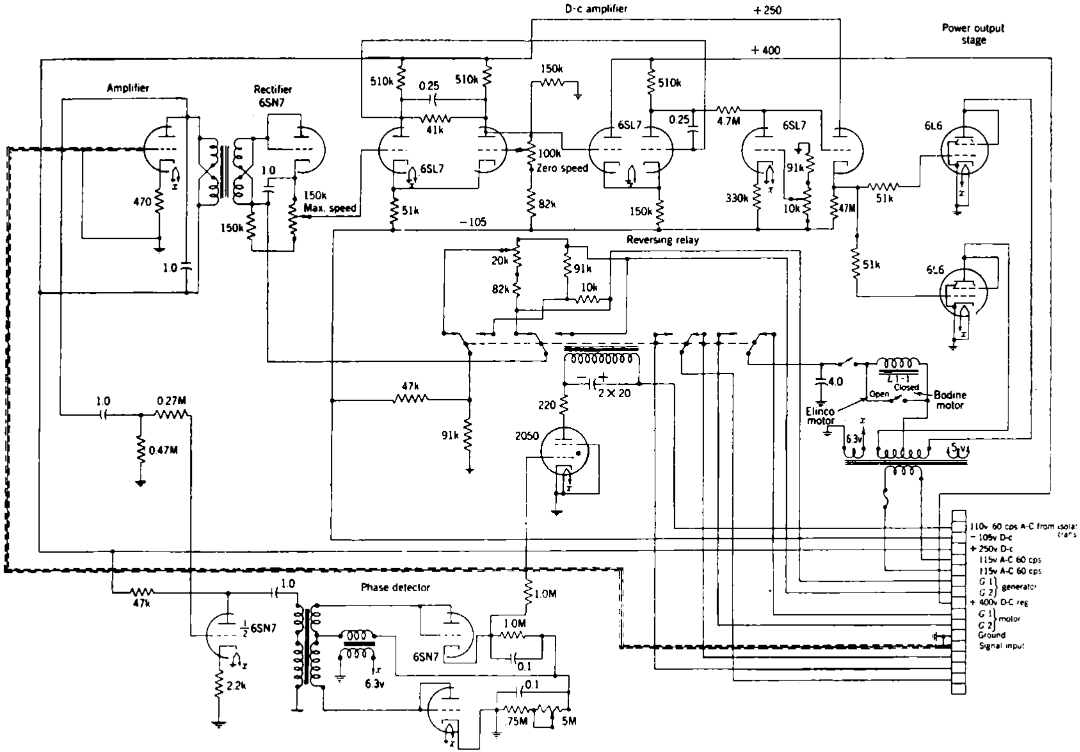


FIG. 7·19.—Velocity servo schematic.

velocity servo input impedance is so high. The respective windings can therefore be connected in series as shown in Fig. 7·13, and no error will exist if the voltages are all in phase.

The output voltages representing the desired vector summations as discussed previously are converted into mechanical shaft rotations whose angular velocities are proportional to the respective voltages by means of the three velocity servos. The three channels are identical. The block diagram is given in Fig. 7·18, and the schematic in Fig. 7·19.

As shown in Fig. 7·18, the 60-cps input voltage to the velocity servo is first amplified and then rectified to obtain a d-c voltage proportional to the amplitude of the input a-c voltage. This d-c voltage is further amplified in a direct-coupled amplifier whose output controls the power developed by the power output stage. The d-c power output is used to

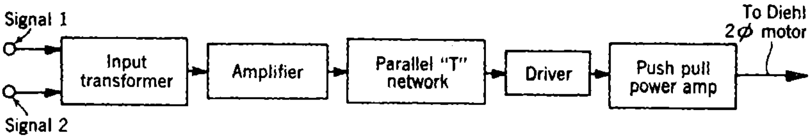


FIG. 7·20.—Servoamplifier block diagram.

drive a Bodine d-c motor after being fed through a reversing relay which is used to control the direction of rotation.

The position of the reversing relay is controlled by a thyratron which in turn receives its control voltage from a phase detector. The phase detector develops a d-c voltage whose polarity and amplitude are dependent upon the phase and amplitude respectively of the input a-c voltage. The thyratron conducts for only a control voltage in phase with the anode voltage. If the input signal reverses, the relay arms change contacts.

From the block diagram (Fig. 7·18) it can be seen that the motor drives a d-c generator whose output voltage is fed back to the rectifier, after feeding through the reversing relay. Thus, when the input phase reverses, the direction of motor rotation reverses which would reverse the polarity of the generator voltage. Since the generator leads are also reversed by the relay, the polarity of the generator feedback voltage is kept the same independent of the phase of the input signal.

The complete schematic diagram for the velocity servo is given in Fig. 7·19. Although no detailed discussion of this circuit is included here, it is fully described in Radiation Laboratory Report No. 645-10.

The block diagram of the servoamplifier that was briefly mentioned in the discussion of Fig. 7·13 is given in Fig. 7·20, while the detailed schematic diagram is shown in Fig. 7·21. The circuit needs very little discussion, since it is very similar to the usual audio power amplifier with the exception of the parallel-T network. This network is included to introduce a variable degree of phase-lead control for stability purposes.

Each of the two input signals is impressed on a primary terminal of the input transformer so that the secondary voltage is proportional to the difference between the two input signals. The difference signal is amplified and applied to the control winding of the two-phase low-inertia servo motor.[1]

Although the general discussion of the ground-range resolver and the three servo loops presented the problems one at a time, this procedure

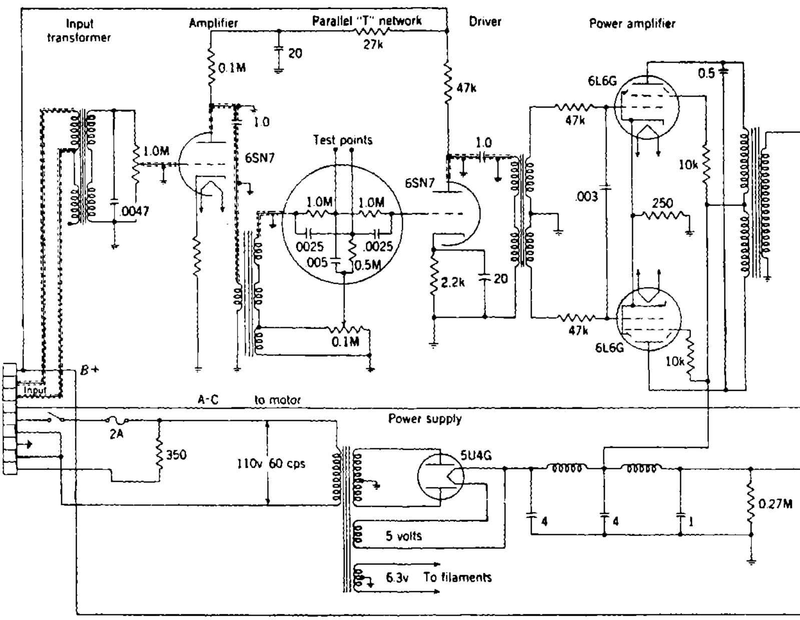


FIG. 7·21.—Servo schematic.

cannot be followed here, since the circuits are more closely interrelated. The circuitry for these channels is given in Fig. 7·22. The element common to all these sections is the range potentiometer that is used to develop an a-c voltage whose amplitude is proportional to the actual slant range.

The potentiometer is fed from a low-voltage tap on a variable auto-transformer which in turn is fed by an isolation transformer from the 60-cps line. Since it is not appreciably loaded, the linearity of the potentiometer (± 0.1 per cent) determines the accuracy by which the displacement of the slant-range shaft is converted to a voltage. The

[1] The operation of this servoamplifier is discussed in RL Report No. 645-2.

1.2-k resistor in series with the range potentiometer serves merely as a voltage divider.

The rotor of the ground-range resolver is driven by a 6AG7 cathode follower whose control grid is directly connected to the arm of the range potentiometer. The cathode-follower circuit is identical to that which is used to feed the rotor of the range rate resolver discussed previously. Since the gain of this cathode follower is nearly constant in the range used, the rotor voltage varies directly as the range and is linear to about ± 0.2 per cent. The resolver output voltages are proportional to the ground range and altitude respectively.

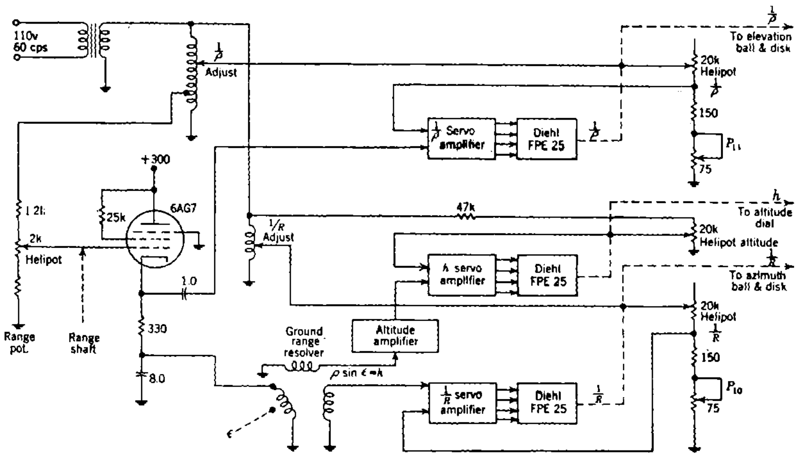


FIG. 7·22.—Ground-range resolver and servo channels.

The range voltage for use in the $1/\rho$ servo loop is also obtained from the cathode follower that feeds the ground-range resolver. The $1/\rho$ potentiometer which is turned by the servomotor is driven by a voltage obtained from the arm of the $1/\rho$ variable autotransformer. The voltage developed across the fixed resistor is fed back to the $1/\rho$ servo input, where it is compared with the actual range voltage. Since the servo loop acts to equalize the two input voltages, the servo motor drives the arm of the $1/\rho$ potentiometer until this condition is fulfilled. The shaft and arm of this potentiometer rotate an amount proportional to $1/\rho$ This same shaft is geared to the ball-positioning mechanism of the elevation ball-and-disk integrator.

The adjustable resistor P_{11} is used to set accurately the value of the proportionality factor between the current flow and the voltage fed back into the $1/\rho$ servoamplifier. The arm of the $1/\rho$ variable autotransformer is used to make the scale factor of the servo matching voltage the same as that of the input range voltage.

The $1/R$ servo loop is identical with the $1/\rho$ loop with the exception that the voltage whose amplitude is proportional to R comes from an output winding of the ground-range resolver rather than a cathode follower. All other circuit details are identical with the channel discussed above.

It has been stated that the actual aircraft altitude is presented as a dial rotation. With an actual radar installation, the altitude must be computed from the other observed data. A voltage whose amplitude is proportional to altitude is obtained from an output winding of the ground-range resolver as has been explained previously. The maximum altitude permitted by mechanical design consideration was set at 35,000

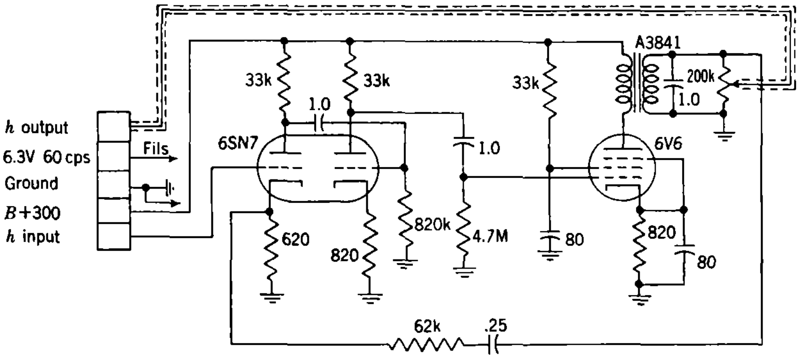


FIG. 7·23.—Altitude amplifier.

ft, whereas the maximum ground range was set at 100 miles. Thus, the maximum amplitude of the altitude voltage from the ground-range resolver is only about 7 per cent of the maximum amplitude of the ground-range voltage. Since a higher voltage level is required for suitable altitude servo operation, a linear amplifier is placed between the altitude output winding and the input to the h servoamplifier. The schematic diagram of this amplifier is given in Fig. 7·23.

A two-stage RC-coupled amplifier is used to drive a 6V6 power amplifier. The output serves as one of the two input signals to the altitude servoamplifier. Since the amplifier must be linear to better than 1 per cent in order that the altitude dial will present information with 1 per cent accuracy, a large value of degenerative feedback is included. The feedback loop returns a portion of the output voltage to the cathode of the first stage where it subtracts from the input signal.

With this amplifier, a gain of 60 is obtained and the linearity is of the order of ± 0.5 per cent. It accepts input voltages up to about 1 volt which corresponds to an altitude of about 40,000 ft. Thus, the amplifier is operated linearly in the desired altitude range from zero to

35,000 ft, and the output voltage level is in the range leading to optimum servo performance.

The matching voltage for the altitude servo is obtained from the arm of the altitude potentiometer which is rotated by the h servo motor. This potentiometer is driven from a line isolation transformer through the 47-k resistor which serves as an attenuator. The servo loop responds in a manner such that the potentiometer voltage is made equal to the h voltage coming from the amplifier. Since the h dial is geared to the altitude potentiometer arm, the dial will be positioned proportional to the potentiometer arm and hence proportional to altitude.

7·11. Over-all System Operation.—Both the theory of operation of the spherical coordinate integrator and the actual form of its reduction to practice have been discussed in considerable detail in the foregoing sections. The way in which the positional data from the integrator is actually presented to the parent radar set in suitable form still remains to be discussed.

The data defining the position of the aircraft with respect to the moving ship ordinarily obtained by the radar set are range, azimuth angle, and elevation angle. The equipment described presents this information in the form of mechanical shaft displacements, as has been shown, but some form of data conversion must be made before this output information can actually be used. In order to understand the necessity for this final data conversion, a brief discussion of radar trainer operation will be found helpful.

The trainer generates i-f pulses corresponding in time to the actual range of the aircraft from the ship. These pulses are fed into the receiver of the parent radar set in place of the i-f signals normally feeding in from the crystal mixer. In order that the proper azimuth and elevation angle information be included, these i-f pulses are gated in accordance with the relative positions of the antenna mount of the radar set and the aircraft. When the position of the two coincide in both azimuth and elevation angle, the i-f pulses feed through; however, if these conditions are not fulfilled, the pulses do not feed through.

The time-modulation circuit that causes the i-f pulses to appear at the proper time following the radar trigger is controlled by a d-c voltage. The integrator must therefore develop a d-c voltage whose amplitude is proportional to the displacement of the range shaft. This voltage can then be used to control the linear delay circuit. This data conversion is performed by gearing the arm of a linear potentiometer to the output range shaft.

The azimuth and elevation angle information is obtained by mounting 360° potentiometers on the radar antenna mount azimuth and elevation shafts and on azimuth and elevation output shafts of the integrator.

Therefore, two azimuth and elevation potentiometers are present; and if they are properly lined up, they can be used in two bridge circuits to furnish the desired gating voltages. The voltage from such a bridge circuit is a minimum when the arms of the two potentiometers in the bridge are in corresponding positions. If either potentiometer shaft is displaced, the output voltage will increase in the usual manner. Thus, if the gating circuits are controlled so that the signal feeds through only when the bridge voltage is a minimum, the synthetic radar echo appears only when both potentiometer shafts are in corresponding positions.

The action is similar with respect to both the azimuth and the elevation bridge circuits, but they respond to motions in planes perpendicular to one another. If the output voltages from the two bridges are properly mixed before being applied to the gating tubes, the gating is dependent upon the coincidence of both the azimuth and elevation angles of the aircraft and the radar antenna mount. The signals, therefore, appear only when the line of sight of the radar antenna mount intersects that of the aircraft.

7·12.—Summary.—Two examples of electromechanical computer design have been presented. While at first glance the devices described appear to be exceedingly complex, upon closer study each is seen to consist merely of a collection of the simple computer circuits and devices presented in preceding chapters (and elsewhere in the Series) coordinated very much along the lines suggested by Chap. 2, and capable of design by straightforward methods. The authors are confident that the near future will see many more computing devices of this same general nature take their places as working tools of science and industry.

The computer designs discussed in this chapter should serve also to underscore a point made earlier; namely, that there is an intimate relationship between computers and servomechanisms. While servomechanisms have been treated as a separate subject in the chapters which immediately follow, it should not be forgotten that these devices are as much a part of the computer designer's "bag of tricks" as any of the devices presented in Part I.

PART II

INSTRUMENT SERVOMECHANISMS

CHAPTER 8

INTRODUCTION AND SUMMARY OF DESIGN PROCEDURE

BY I. A. GREENWOOD, JR.

INTRODUCTION

8·1. General Principles of Servomechanisms.—The design and use of servomechanisms have grown to be an extremely important part of electronic and mechanical technology. The increasing demands of engineering and science for greater accuracies, speeds, and efficiencies; the comparative newness of many aspects of the subject; and the increasing availability and use of wartime developments are factors that tend to make the servomechanism field one of rapid growth and widespread interest at this time. The exacting technical requirements of military devices brought a great acceleration in servo development during World War II, and servos for military purposes were produced in vast quantities during the war.

For the purposes of this book the definition of the term "servomechanism" as proposed by Hazen[1] and used by Hall[2] and others will be used. According to this definition a servomechanism is "a power-amplifying device in which the amplifying element driving the output is actuated by the difference between the input and the output." An example of a servomechanism is the simple data-transmission system of Fig. 8·1. With this system it is possible to turn the input shaft through the angle θ_i and to have this motion repeated by an output shaft rotation of θ_o at a remote location, with a power amplification. If the rotation of the output θ_o is different from the rotation of the input θ_i, an error voltage e is developed across the rotor leads of the synchro control transformer.[3] This error voltage is phase- or sense-detected to yield a d-c signal that when amplified and used to control power to the motor will result in rotation of the motor tending to make θ_o correspond to θ_i. Other schemes accomplishing the same end are, of course, used. At first glance, it would appear that this is a fairly simple and potentially

[1] H. L. Hazen, "Theory of Servomechanisms," *Jour. Franklin Inst.*, **218**, No. 3, 279–330, September 1934.

[2] A. C. Hall, *The Analysis and Synthesis of Linear Servomechanisms*, MIT Servomechanisms Laboratory, 1943. (Reprint of MIT doctorate thesis.)

[3] See Vol. 17 for a detailed discussion of control transformers. The subject is also treated in Sec. 13·2.

very useful device. This is true. Its apparent simplicity, however, may be misleading in that careful and intelligent design based on a knowledge of feedback theory is needed in order to ensure that the resulting simple device will operate as desired. There may be a surprisingly close resemblance between the circuit diagram of a useful and well-behaved servomechanism and the circuit diagram of a servomechanism that will burst into violent oscillation the moment power is applied. Once the dangers are recognized, however, the techniques of servo design are sufficiently straightforward so that design and use of servos can and

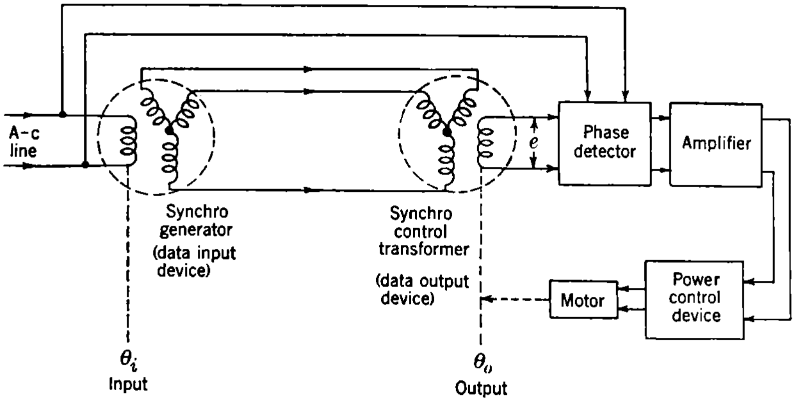


FIG. 8.1.—Simple remote follow-up servomechanism.

should become a useful professional tool for the average engineer and scientist.

It is of interest to inquire further into the advantages that servomechanisms may offer. An important advantage follows from the definition, that is, the ability to amplify power, usually mechanical power. A second advantage, mentioned in connection with the servo of Fig. 8·1, is the ability to transmit information from one place to another. The combination of these two advantages, the ability to control large powers remotely, has accounted for the development of a large fraction of the servos that have been used to date. An advantage somewhat related to the second listed above is the elimination of reaction on the controlling element when large powers are controlled. An interesting application in which this property is of importance is the use of servomechanisms to allow unilateral flow of torque from one element to another in the M.I.T. differential analyzer.[1] Other advantages of servomechanisms are the result of their ability to effect transformations from one type of data

[1] Bush and Caldwell, "A New Type of Differential Analyzer," *Jour. Franklin Inst.*, **240**, 255–326, October 1945.

representation to another, the most common transformation being from electrical data to mechanical data. This ability to transform from one data representation to another is of itself a valuable feature; it also allows the techniques and advantages of feedback to be applied to systems involving other than purely electronic elements (e.g., mechanical and electromechanical elements), thereby increasing their accuracy of operation, speed, and efficiency and making possible automatic operation.

In Fig. 2·7 a servomechanism is used to operate the bridge computer of Fig. 1·1. This illustrates the use of a servomechanism to apply feedback techniques to a loop involving electromechanical elements, thereby changing the bridge into a device that may operate automatically.

8·2. Uses of Servos.—All of the advantages of servomechanisms mentioned in the previous section apply to servomechanisms used in computers. This type of usage was discussed in Sec. 2·7. Examples of computers using servos are numerous, particularly in the field of military devices. A differential analyzer application has already been mentioned. Servos have been used in computers for ground- and ship-based antiaircraft gunnery, air-to-air gunnery, and bombing.

The ability of servos to facilitate control at a distance finds many industrial and military applications. Practically all heavy antiaircraft artillery, for example, is servo-controlled by the output of some type of computer usually located remotely from the guns. The use of servos for control of industrial machinery is a rapidly expanding field, a spectacular example of such an application being a servo-controlled aircraft wing spar milling machine[1] which cut the time required for a complicated machining job from $13\frac{1}{2}$ hr to 5 min. Servomechanisms have been used for turning the tuning condensers of push-button-controlled radios. An interesting application of servomechanisms has been in the manufacture of fissionable material[2] for the atomic bomb, in which whole complicated processes were remotely controlled. This is but one example from what is probably the largest field of application of servomechanisms—industrial process control. An extensive technology[3] on process control has already been built up. It is of interest to note that many of the control devices used in present-day process control techniques are based on other than electronic methods.[4] There is good reason to believe that many of the electronic devices reviewed in this volume and similar

[1] *Electronics* **17**, 146, October 1944.

[2] H. Smyth, *Atomic Energy for Military Purposes*, U.S. War Dept., 1945, Secs. 7, 27, and Appendix 4.

[3] See for example, D. P. Eckman, *Industrial Process Control*, John Wiley & Sons, New York, 1945; and E. S. Smith, *Automatic Control Engineering*, McGraw-Hill Book Co., Inc., New York, 1944.

[4] *Cf.* Sec. 12·15.

devices will see greatly increased use in the process control field during the next few years.

8·3. Definitions of Terms and Concepts.—Hazen's[1] and Hall's[2] definitions of the term servomechanism as "a power-amplifying device in which the amplifying element driving the output is actuated by the difference between the input and the output" has already been mentioned; for the purpose of this book, this definition will be used. There are alternate definitions in the literature, however, that are of interest. The ASME *Proposed Glossary of Automatic Control Terms*[3] uses the term "automatic controller" in nearly the same sense that servomechanism is here used, defining it as a "mechanism which measures the value of a variable quantity or condition and operates to correct or limit deviation of this measured value from a selected reference." Two differences are noted: lack of an implied power amplification and use of a "selected reference" as opposed to "an input."

Current usage considers the words "servo" and "servomechanism" as equivalent, although Brown gives a distinction between the two terms in an early paper.[4] A recently proposed British definition[5] of the term "servo system" is "a power amplifying, automatic, error-actuated control system."

Following a definition by Harris,[6] a "regulator" is considered to be a special type of servomechanism which tends to keep a physical quantity at a constant level, whereas a servomechanism may make the physical quantity vary over a predetermined cycle or vary as a definite function of some other arbitrarily varying quantity. A recently proposed British definition[7] of the term "automatic regulator system" differs only slightly from Harris' usage in defining it as "an automatic error-actuated control system, the input signal to which is preset to a constant value or to a series of values varying with respect to time in a predetermined manner." The British[8] definition will be used in this volume for the term "automatic control system" or merely "control system," defined as "an arrangement

[1] Hazen, *loc. cit.*

[2] Hall, *loc. cit.*

[3] Reproduced in Eckman, *op. cit.*, pp. 224–230. ASME stands for American Society of Mechanical Engineers.

[4] G. S. Brown, "Behavior and Design of Servomechanisms," NDRC Sec. D-2 Report, November 1940.

[5] Ministry of Supply Servomechanisms Panel, *Glossary of Terms Used in Control Systems with Particular Reference to Servomechanisms*, published by Military College of Science, January 1946.

[6] H. Harris, "The Analysis and Design of Servomechanisms," NDRC Report, 1942.

[7] Ministry of Supply Servomechanisms Panel, *op. cit.*

[8] *Ibid.*

of elements interconnected in such a way that the operation of each depends on the result of the operation of one or more other elements, the purpose of which is to control some condition of a body, process, or machine." Hall's[1] usage of this term is nearly the same but includes the idea of power amplification.

It is of importance to distinguish between the types of control that Hall has referred to as "open-cycle control" and "closed-cycle control." Open-cycle control works on signals received solely from a controlling instrument, while in closed-cycle control system, additional signals that are derived from the position of the device or the state of the process being controlled are received by the controller. Hall's term "open-cycle control" has been referred to by the ASME *Glossary*[2] as "some form of automatic operation," and by the British as an "unmonitored control system" or "input-actuated control system."

The term "error" is used by most authors to represent the difference between the servo input and output. The ASME recommended term for this is "deviation." The terms "misalignment" and "difference" are occasionally used.

When highest accuracy in a servomechanism is required, a "continuous control system" is usually used. Hall[3] defines such a system as one in which "a definite and continuous corrective action is developed by the servo controller and applied to the device being controlled no matter how small is the error in the position of that device." The majority of the devices covered in Part II of this volume will be continuous-control devices or devices operated such as to closely approximate continuous control. For example, a vacuum-tube motor control is a "continuous control"; an on-off relay motor control is not.

For the purpose of this volume, a distinction must be made between instrument servos and power servos. The present treatment is limited to instrument servos. An instrument servo is arbitrarily defined as a servo rated at less than 100 watts maximum continuous output. The term "instrument" in this title is derived from the fact that most of the servos discussed will be those used in instruments, but exceptions will be found. A synchro system by itself is not here considered to be a servo, since it involves no power amplification, and is not treated. The use of straight synchro systems is discussed in Vol. 17, while Vol. 25 treats higher-power servomechanisms.

The term "nonlinear servo" will be used frequently in this volume. Two types of nonlinearity are recognized. In one type, elements such as amplifiers or motors are nonlinear. In the other type, the gain of the

[1] Hall, *loc. cit.*
[2] Eckman, *op. cit.*, pp. 224–230.
[3] Hall, *loc. cit.*

servo varies over the working range of the inputs and outputs; but for any small region of the working range, the performance approximates that of a servo with linear elements operated at a gain determined by the particular region of operation chosen. An example of this type of nonlinearity is the resolver servo of Sec. 14·3. Obviously, if gain varies drastically over a small region, nonlinearity equivalent to the first type may result. Both types of nonlinearity may be present at once. Nonlinearity is treated in Sec. 11·11.

A convenient classification of servos may be made in accordance with their uses, the principal examples of which are "position servos" and "velocity servos." A "position servo" is one in which the displacement of an output shaft or its equivalent is controlled by some input variable. The simple servo shown in Fig. 8·1 is thus a position servo. A "velocity servo" is one in which the first derivative with respect to time of the output is controlled by the input variable or the error signal. In some cases, there may be difficulty in deciding which of these classifications is appropriate. Consider, for example, the case of a velocity servo used to integrate an input voltage. Since the input voltage is usually measured across a resistance, the input might be thought of as charge rather than voltage, in which case the output rather than the derivative of the output would seem to be the quantity directly controlled by the input, and one would be tempted to classify this as a position servo. Its correct classification as a velocity servo is clarified by examination of the error signal controlling the output. In this case the error signal is a voltage that is a function of the first derivatives of assumed input and output. There are a few types of servos that do not fit either position or velocity servo definitions very well; nevertheless, these descriptive terms are widely used and are appropriate in most instances. Servos could also be designed such that the second or higher derivatives of the output would be controlled by some input variable, but such servos are extremely rare.

8·4. Plan and Scope of Part II.—Following the above introductory sections are a number of sections dealing with the recommended procedures and techniques for designing electronic instrument servo systems. These constitute a summary only, with references to following sections where detailed discussions of each of the various components may be found. Chapters 9, 10, and 11 present a theoretical background for the study of the characteristics of servo components and of over-all servo systems, emphasizing particularly factors of stability and over-all accuracy. This theoretical discussion approaches the servo design problem from both the transient and steady-state viewpoints. A number of special problems, such as data smoothing, nonlinearity, gear ratios, etc., are discussed in Chap. 11. Chapter 12 is devoted to a rather detailed discussion of the principal components of the usual electronic servo systems, namely: data input and output devices; amplifiers, phase detectors, modulators, etc.; motors and other power devices; and power-control circuits. Chapter 13 summarizes practical techniques for obtaining measurements of the characteristics of servo components and servo systems. Chapter 14 summarizes a number of special servo systems. Detailed circuit diagrams are included.

Although the emphasis of this treatment is on electronic devices, brief references to some competing nonelectronic devices have been included.

DESIGN TECHNIQUES

8·5. Preliminary Design Data.—The first step in the design of a servomechanism is the *determination of what must be designed.* This is important, and time spent in carefully stating design requirements will generally be saved in succeeding stages of the design.

Type of Servo Required.—It must first be decided what functional type of servo is required, that is, velocity servo, position servo, etc. Other classifications of servos that have been mentioned, such as continuous vs. noncontinuous or linear vs. nonlinear, specify alternate ways of fulfilling given requirements rather than fundamental classifications of the requirements with which this section is concerned.

Data Input and Output Representations.—The input and output representations of data must be specified. The subject of data representation is discussed in Secs. 2·11 to 2·13 and in Secs. 12·1 to 12·6. Before the servo of Fig. 8·1 could be designed, for example, it would be necessary to know that the input data representation is a mechanical shaft rotation and that the output data representation is also a shaft rotation. Scale factors and ranges must, of course, also be known.

Power Available.—The power available should be specified as to voltages, allowable currents, and, if alternating current, the frequency. Voltage and frequency tolerances are very important in servo design and should be carefully determined.

Quality.—It is important to specify accurately the minimum acceptable quality. A design procedure must attempt to yield a servo a factor of safety better than this minimum; to go very much further may be a waste of expensive equipment and design time. In many cases the requirements on a servo are so severe that one must go to extremes of present techniques to achieve satisfactory performance; in other cases cheap simple devices and rudimentary design may suffice.

Quality may be expressed in a variety of ways. Maximum, average, or probable errors may be stated for specified inputs. Probable or average errors are usually associated with input and noise specified in terms of their power-frequency distributions. Maximum errors under the special input conditions of velocity and position steps of given magnitude are frequently used. Steady-state velocity and accleration errors are very convenient means for specifying quality; it will be shown later that they may easily be related to the feedback loop decibel gain vs. log frequency characteristics of the servo. Bandwidth is frequently used as an index of "speed of response," a term somewhat loosely used but referring principally to the acceleration of a servo.

It is usually desirable to specify the damping of the servo. For this the damping factor (or factors) may be used; a more convenient index of stability is the height of the peak of the sinusoidal steady-state over-all response curve (*cf.* Chaps. 9 to 11). The smoothness of operation may also be specified.

Load Specification.—It is necessary to specify the force or torque and inertial load on the output of the servo and the speed range over which it must operate, including slewing.

Other Factors.—The list of design factors of Chap. 19 should be checked; appropriate factors listed as design requirements; and specific limits and operating conditions chosen for these design factors. Factors applying particularly to servos are the following: slip ring and commutator electrical noise, operating position, backlash, hysteresis, life, and reliability.

8·6. Design Procedure.—The following discussion of design procedure is intended merely to summarize the recommended steps in designing a servomechanism after the preliminary design data have been assembled. For most steps, a reference is given to a succeeding section for detailed treatment of the problems involved. Although arbitrary, the procedure has been found to be useful.

After completing the preliminary specifications, the designer may proceed as follows:

1. Choose the motor and motor-control circuit or their equivalents and the approximate motor gear reduction. Chapter 12 treats these components. Problems of gear ratios, friction, and backlash are discussed in Sec. 11·12.
2. Choose the data input and output devices and their approximate gearing, unless these have already been specified in the preliminary design data. See Secs. 12·1 to 12·8 for a detailed discussion of data input and output devices.
3. Choose types of circuits for the amplifiers, phase detectors, modulators, demodulators, etc., required. In the choice of these circuits sufficient amplification should usually be allowed to make up a loss in gain of roughly 10 (20 db) due to phase-lead circuits which

may be used for stability in addition to the usual factor of 2 to 4 for tube aging, line voltage changes, component aging, etc. See Secs. (12·10 to 12·12 for further discussions of these circuits. An important part of this design step is a choice of the type or combination of types of controller characteristics that are to be achieved. Chapters 9 to 11 discuss these characteristics.

4. Make a first design in some detail for each of the above portions of the servomechanism. It should be kept in mind that only parts procurable in the desired quantities should be specified, even in this early phase of the design. The remarks of Sec. 2·2 apply here.
5. Determine by calculations, measurements, and reference to previously established data the detailed characteristics of the motor and control equipment; the amplifiers, phase detectors, modulators, etc.; and the input and output devices. This may or may not include phase-lead circuits or devices. See Chap. 13 for a discussion of the techniques of such measurements. The theory of Chaps. 9 to 11 will be found helpful in any calculations required.

   It may be desirable to build up a breadboard model of the power-control circuit so that empirical data on the motor and control circuit combined may be obtained. This procedure is particularly justified when local feedback is used in the power-control and motor circuits, resulting in more nearly linear over-all characteristics for the combination than for the motor alone or when the waveforms of the controlling circuit are so complex as to make the accuracy of theoretical calculations questionable or the labor of computation excessive.
6. On the basis of the information obtained in the last step, design in detail the phase-lead circuits, integral control circuits, amplifiers, velocity feedback circuits, dampers, or other devices or circuits where changes in the shape of the transmission-frequency characteristics will improve the stability, accuracy, or smoothness. The theoretical treatment of Chaps. 9 to 11 should be referred to.
7. Work over the design to make all elements consistent with the preliminary design considerations and specifications and with each other.
8. Build a prototype or breadboard model to test performance and as a check on the theoretical calculations. It is good practice to test such a model at some time with all limit tolerance parts, with tolerances chosen such that resulting errors will add.

   A breadboard model or prototype model also allows the final adjustment of component values that is often necessary even with good theoretical design.

9. Determine tolerances and specifications for each and every component. A design cannot be considered complete until this has been done in every detail and the component is shown to be obtainable in the quantity required. Failure to complete this may mean that a single unsatisfactory component will necessitate a complete redesign, although this is not likely.

The steps of this procedure, particularly the first three, are intimately related. *It will usually be necessary to repeat the procedure several times before a fully satisfactory design is achieved.*

8·7. Design of Servos by Experimental Techniques.—For low-quality servos or for designs that are similar to previously tested designs, it is sometimes safe to omit the theoretical design of stabilizing circuits. Final component values under such a shortened procedure are determined by experiment and test, using limit tubes and components if possible. Preliminary design is based on sufficient gain to allow for a stabilization circuit attenuation factor of roughly 10, unless other stabilizing means are used, and must include the right type of controller characteristics as mentioned in Step 1 of the design procedure of the preceding section. *It is important that the servo designer understand the theory of use of these various types of controller circuits regardless of whether or not the theoretical calculations are made.* A person who works with servos can rapidly acquire an ability to adjust an existing design to near-maximum performance. When this ability is combined with a basic understanding of the theory of servomechanisms and of the many special-purpose circuits and devices available, the resulting combination of skills will frequently allow good servos to be produced rapidly without extensive theoretical calculations. Acquisition of the adjustment skill alone, however, is usually wasteful of time in the long run.

Where a design is to be produced in quantity, it is desirable to know what the accuracy and stability safety factors will be under adverse combinations of production component tolerances. The extreme importance of having this information for production designs normally justifies both theoretical analysis and careful experimentation, although careful experimentation alone may suffice in some cases.

CHAPTER 9

SERVO THEORY: INTRODUCTION AND TRANSIENT ANALYSIS

By G. L. Kreezer

INTRODUCTION

9·1. The Aims of Servo Theory.—The following three chapters deal with the application of mathematical methods to problems of servo design and adjustment. A later chapter deals with the considerations relevant to the selection of the components of a system intended to meet given performance specifications. Ways of evaluating this provisional choice to determine if a system made up of these components will, in fact, perform as required will be considered here. Theoretical principles are introduced at this point, since it will be helpful for them to be kept in mind in the survey of available components that follows.

To evaluate a system made up of a given set of components, the most obvious procedure is to construct a trial model and determine its performance through actual observation and test. Such a procedure is more readily feasible in the case of small instrument-type servos than it is for systems of greater power level, but even here it may be uneconomical of time and materials, requiring a long sequence of trial constructions and tests. It is natural, therefore, to attempt to carry out the trial construction on paper through the medium of a mathematical model.[1]

So-called servo theory makes possible the construction of a symbolic model of the proposed system in the form of mathematical equations, and the carrying out of the appropriate tests, experiments, and adjustments on this model by means of mathematical operations. The basic questions in servo theory are, therefore, the same obvious ones that any experiments with a physical model would attempt to answer.

1. How does this system perform?
2. How does its performance compare with the standard specifications, set up on the basis of practical needs?
3. If it fails to meet performance specifications, how can it be modified so that it will do so?

[1] Theoretical treatment has the additional merit of facilitating specification of production tolerances of the components to be used in a given system. For on a mathematical basis, the effect on performance of variations in magnitude of component parameters may be predicted. Cn an experimental basis, trials are necessary of the worst cases to be expected in practice.

The methods of servo theory are designed merely to make possible answers to these simple questions. In the present account, these questions will be used to provide the framework for our survey. Thus, under the subject of the *determination of system performance* will be considered the methods available for finding out how a given system performs; under the subject of *evaluation of system performance*, the ways available for comparing its performance with specifications; and under the subject of *correction of system performance*, ways for removing deficiencies in performance through suitable modifications in the structure of the system. In a final section on *special problems*, some problems of a more complex or special type will be examined. But it is necessary to review first certain preliminary concepts and definitions that will be utilized in subsequent discussions.

PRELIMINARY CONCEPTS AND METHODS

9·2. Transformation and Operational Methods.—In the present survey of servo theory, considerable use will be made of the Laplace transformation. It is necessary, therefore, to review some of the salient features of the method.

The Laplace transformation provides a way of representing a function of a real variable $f(t)$ by a function of a complex variable $F(s)$, and conversely. The variable t will be regarded here as standing for time, and the complex variable s as standing for $\sigma + j\omega$. By means of the *direct* Laplace transformation, we may pass from the real function $f(t)$ to the complex function $F(s)$. This transformation is designated symbolically as

$$\mathfrak{L}[f(t)] = F(s)$$

and is read "the Laplace transform of $f(t)$ equals $F(s)$." By means of the inverse Laplace transformation, we may pass from $F(s)$ to $f(t)$. In symbols,

$$\mathfrak{L}^{-1}[F(s)] = f(t)$$

and is read, "the inverse Laplace transformation of $F(s)$ equals $f(t)$."[1] The equations[2] defining these two transformations are

[1] In representation of functions in the two domains, the practice of Gardner and Barnes of representing functions of a real variable by small letters and the corresponding function of a complex variable by large letters will, for the most part, be followed. Occasionally, the same letter will be retained for the corresponding functions in the real and complex domains, where ambiguity might otherwise result. In every case, however, the domain in which the function lies will be indicated by the variable inside the parenthesis. Thus $e_1(t)$ is a function of time, and $e_1(s)$ is its Laplace transform, a function of the complex variable s.

[2] For a discussion of these equations and the conditions for their validity see

$$F(s) = \mathfrak{L}[f(t)] = \int_0^\infty f(t)e^{-st}\,dt, \qquad \sigma_a < \sigma, \tag{1}$$

and

$$f(t) = \mathfrak{L}^{-1}[F(s)] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F(s)e^{st}\,ds, \qquad \begin{matrix} 0 \leqq t \\ \sigma_a < c \end{matrix} \tag{2}$$

The possibility of a function being represented as a function either of a real variable or of a complex variable is of value in that a problem difficult to solve in one domain may be relatively easy in the other. Difficulties may then be bypassed by shifting to the less difficult domain. In the case of the problems arising in servo theory, this advantage will be utilized chiefly in the provision of relatively simple procedures for the solution of linear differential equations. Differential equations involving functions of time become after transformation algebraic equations in s. These equations can be manipulated easily by purely algebraic operations. We may solve for the function of interest, in the complex domain, and then pass to the corresponding real function that provides the solution of our differential equation. Certain special advantages arise in relation to the concept of transfer function which is considered below. It should be noted that since our direct transform $F(s)$ is a function of a complex variable, it is feasible to apply, where useful, any of the special procedures that form a part of complex function theory.

So much by way of preliminary orientation. The chief relationships that we shall need to use in our treatment of servo systems may now be summarized.[1] The starting point for a given problem will typically be a differential equation made up of functions of t. The first step in the use of the Laplace transformation will consist in application of the direct Laplace transformation to both sides of the equation. This operation might be carried out by application of the defining Eq. (1). But since certain types of function and algebraic operations on functions repeatedly occur in different equations, it has been found convenient to set up tables of transform pairs such as Tables 9·1a and 9·1b, which

---

M. F. Gardner and J. L. Barnes, *Transients in Linear Systems*, Vol. I, Wiley, New York, 1942, pp. 100–107.

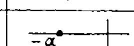
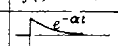
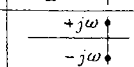
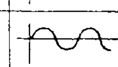
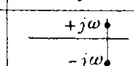
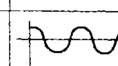
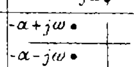
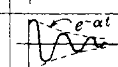
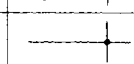
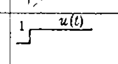
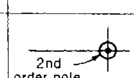
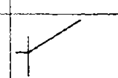
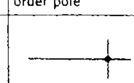
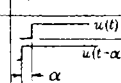
[1] For the derivation of these relationships and a fuller treatment of the Laplace transformation, the following references may be consulted: Gardner and Barnes, *loc. cit.;* N. W. McLachlan, *Complex Variable and Operational Calculus*, Cambridge, London, 1942; R. V. Churchill, *Modern Operational Mathematics in Engineering*, McGraw-Hill, New York, 1944; G. Doetsch, *Theorie und Anwendung der Laplace Transformation*, Springer, Berlin, 1937; E. S. Smith, *Automatic Control Engineering*, McGraw-Hill, New York, 1944, pp. 302–308. A brief treatment may be found in Vol. 18 of this series. The present account is based largely on Gardner and Barnes.

show the result of applying the direct and inverse transformations to commonly occurring functions and operations.

TABLE 9·1*a*.—LAPLACE TRANSFORM PAIRS FOR OPERATIONS

No.	$f(t)$	$F(s)$	Special conditions and definitions
a	$af(t)$	$aF(s)$	a is a constant or a variable independent of t and s
b	$f_1(t) \pm f_2(t)$	$F_1(s) \pm F_2(s)$	
c	$\dfrac{df(t)}{dt}$	$sF(s) - f(0+)$	$f(0+)$ is value of $f(t)$ immediately following $t = 0$
d	$\dfrac{d^2f(t)}{dt^2}$	$s^2F(s) - sf(0+) - f'(0)$	$f'(0) = \left.\dfrac{df(t)}{dt}\right]_{t=0}$
e	$\int f(t)\,dt$	$\dfrac{F(s)}{s} - \dfrac{f^{-1}(0+)}{s}$	$f^{-1}(t) = \int f(t)\,dt$ $f^{-1}(0+) = f^{-1}(t)]_{t=0+}$

TABLE 9·1*b*. LAPLACE TRANSFORM PAIRS FOR FUNCTIONS*

Pair No.	$f(t)=\mathfrak{L}^{-1}[F(s)]$	$F(s)=\mathfrak{L}[f(t)]$	σ_a	Location of poles	Type of $f(t)$ function
a	$e^{-\alpha t}$	$\dfrac{1}{s+\alpha}$	$-\alpha$		
b	$\sin \omega t$	$\dfrac{\omega}{s^2+\omega^2}$	0		
c	$\cos \omega t$	$\dfrac{s}{s^2+\omega^2}$	0		
d	$e^{-\alpha t}\cos \omega t$	$\dfrac{s+\alpha}{(s+\alpha)^2+\omega^2}$	$-\alpha$		
e	1, or $u(t)$	$\dfrac{1}{s}$	0		
f	t	$\dfrac{1}{s^2}$	0		
g	$u(t-\alpha)$	$\dfrac{1}{s}e^{-\alpha s}$	0		

* Modified from tables in Gardner and Barnes, *op. cit.*, pp. 120, 332, 356: and E. S. Smith, *Automatic Control Engineering*, McGraw-Hill, 1944, pp. 322*ff*.

Table 9·1*a* shows the effect of applying the direct transformation to the most commonly occurring *operations*. In Column 2 is shown the real function $f(t)$; in Column 3 the equivalent complex function $F(s)$, obtained after application of the direct transformation. Column 4 specifies the special conditions under which the correspondence is valid.

Table 9·1*b* is a similar table for specific *functions*. Column 2 gives the real function $f(t)$; Column 3 the corresponding complex function $F(s)$. In any given problem, one need only find the appropriate function in the table and can then write down the function in the other domain to which it corresponds, much as he would use a table of integrals. Column 4 gives σ_a, the σ of absolute convergence, i.e., the minimum value of the real part of s for which $f(t)$ can be regarded as the inverse transform of $F(s)$.[1] Columns 5 and 6 give a graphical representation of the transform pair in question. Thus column 5 shows the location in the complex plane of the poles of the function $F(s)$, and Column 6 the graph of the type of time function corresponding to this location of the poles.[2] Determining the location of the poles of a complex function $F(s)$ is of great importance, since the location and order of the poles establish the nature of the time function that appears on carrying out the inverse transformation.

It may be of interest to add a few words on the relation of the Laplace transformation to the historically older operational methods derived from Heaviside. The Laplace transformation method has been characterized as the modern equivalent of the Heaviside operational calculus. The expressions derived by the two methods show marked similarities. Smith has aptly referred to them as "two dialects of the same mother tongue."[3] The chief differences consist in the method used to represent input functions and in the method of inserting initial conditions. The differences in appearance are at a minimum when the initial conditions are zero. Thus, if with initial conditions zero the Laplace transformation is applied to a differential equation to obtain an algebraic equation in s, the same result could be obtained by substituting s for the differential operator d/dt and $1/s$ for the integration operator, just as in the case of the operational calculus. Where time functions such as $\theta_i(t)$, $\theta_o(t)$, or $E(t)$ occur, they are still written as functions of t in the case of the

[1] See Gardner and Barnes, *op. cit.*, pp. 102, 122*ff*.

[2] It will be recalled that the poles of a complex function $F(s)$ constitute a certain type of singularity of the function, singularities being the values of s for which $F(s)$ or its first derivative fails to be finite and single-valued. (See McLachlan, *op. cit.*, pp. 8–14.) If it is assumed that $F(s)$ is a rational function, it may be represented as the ratio of two polynomials $P(s)$ and $Q(s)$ of degrees m and n respectively. If the roots of $P(s)$ and $Q(s)$ are known, then the function $F(s)$ can be represented as in Eq. (3).

$$F(s) = \frac{P(s)}{Q(s)} = \frac{(s + s_a)(s + s_b) \cdots (s + s_m)}{(s + s_1)(s + s_2) \cdots (s + s_n)}, \tag{3}$$

where $-s_a$, $-s_b$, . . . , $-s_m$ are the roots of $P(s) = 0$, and $-s_1$, $-s_2$, ... , $-s_n$ are the roots of $Q(s) = 0$. If identical factors in numerator and denominator have been canceled out, and if s is given the value of any one of the roots of the denominator, $Q(s)$ will equal zero and $F(s)$ will equal infinity. The roots of $Q(s)$ are therefore designated as "poles," first-order poles if the root occurs only once, second-order poles if a particular root occurs twice, and so on. In similar manner, the roots of $P(s)$ are called the "zeros" of the function $F(s)$.

[3] Smith, *op. cit.*, p. 303.

Heaviside operational method but are written as functions of s in the case of the transformation method. Consequently, in the operational calculus, operational expressions in p may be found combined with time functions, whereas such hybrids will not occur if one is using a transformation method. Equations will contain only functions of time in the real domain or only functions of s in the complex domain. A more fundamental difference between the two methods lies in the fact that the s of the expressions obtained by means of the Laplace transformation is a complex variable whereas the p or $1/p$ of the expressions obtained in the Heaviside calculus are merely symbols for the operations of differentiation and integration. Despite this difference in the fundamental significance of the variable s or p,[1] it has been customary to refer to both types of expression as operational expressions, a practice that will also be followed here.[2]

9·3. Transfer Functions.—A concept that pervades the entire field of servo theory and is basic in the application of the frequency method of analysis is that of *transfer function*.[3] The nature of the concept may be understood by considering its relation to the differential equation used to describe a given system. Let us assume a system describable by an ordinary differential equation with constant coefficients, such as Eq. (4). The independent variable is time t, and the dependent variables θ_i and θ_o are regarded as functions of time. In functional notation they may be written $\theta_i(t)$ and $\theta_o(t)$. The term $\theta_i(t)$ can be regarded as the forcing function or input signal of the system, and $\theta_o(t)$ as the response or output signal. The transfer function is defined as the ratio of the output to

[1] The particular symbol used in operational expressions whether s, p, λ, or any other is, of course, trivial.

[2] For further discussion of the relation of different operational methods, see Gardner and Barnes, *op. cit.*, pp. 99–107 and pp. 359–366; McLachlan, *op. cit.*, pp. vi, 115; and Smith, *op. cit.*, pp. 302–306.

[3] Somewhat different names have been used by different authors in referring to this concept. Thus Gardner and Barnes, *op. cit.*, pp. 132, 152, use the term *system function*. This is a term which we should have preferred except for the possible implication here that the term applies only to the entire servo system. H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, pp. 15, 227*ff*., in general uses the term *immittance function*, although at times he also uses the term *transfer function* in the sense used here. A. C. Hall, *Analysis and Synthesis of Linear Servomechanisms*, MIT, uses the term *transfer function;* but in defining it, he seems to limit it to what we call the *feedback transfer function*. Writers following the terminology of the Heaviside operational calculus use the term *operators*. McColl, *Servomechanisms*, Van Nostrand, New York, in dealing with steady-state relations, uses the term *transmission ratio*. The familiar concepts of transfer impedance and admittance functions of electrical circuit theory may be regarded as special forms of transfer function. See E. A. Guillemin, *Communication Networks*, Vol. 2, Wiley, New York, 1935, p. 475.

the input, after the differential equation has been transformed to an algebraic equation through the application of some one of the methods of operational calculus (Heaviside calculus, Fourier transformation, or Laplace transformation). These steps are carried through for Eq. (4), with the Laplace transformation being used to transform the equation. We may assume that the initial conditions are all zero.

$$J\,\frac{d^2\theta_o(t)}{dt^2} + f\,\frac{d\theta_o(t)}{dt} = \theta_i(t), \tag{4}$$

$$\mathcal{L}\left[J\,\frac{d^2\theta_o(t)}{dt^2}\right] + \mathcal{L}\left[f\,\frac{d\theta_o(t)}{dt}\right] = \mathcal{L}[\theta_i(t)], \tag{5}$$

$$Js^2\theta_o(s) + fs\theta_o(s) = \theta_i(s), \tag{6}$$

$$(Js^2 + fs)\,\theta_o(s) = \theta_i(s), \tag{7}$$

and

$$\frac{\theta_o(s)}{\theta_i(s)} = \frac{1}{Js^2 + fs}. \tag{8}$$

In accordance with our definition, the ratio $\theta_o(s)/\theta_i(s)$ may be designated as the transfer function of the system. The chief advantages resulting

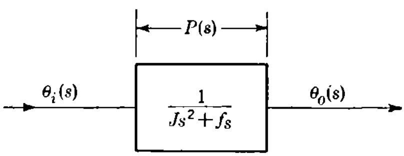


FIG. 9·1.—Block diagram illustrating use of transfer function.

from use of the transfer function concept in servo theory may be briefly summarized.

1. The transfer function fits in readily with the representation of complex systems by means of block diagrams. Thus a given block, in a block diagram, corresponds to the transfer function. The forcing function, in operational form, is considered as the input; and the response, in operational form, as the output. Thus Eq. (8) may be represented by the block diagram of Fig. 9·1. A sequence of such units in cascade, as in Fig. 9·2, can be used to represent a set of independent transfer functions. This correspondence will be clearer from the discussion below on the combination of transfer functions.

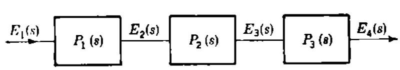


FIG. 9·2.—Combinations of transfer functions.

2. The response of a given system to an arbitrary forcing function can be simply represented as the product of the forcing function, in operational form, and the transfer function. This result follows directly from the definition. Let us represent the transfer function as $P(s)$.

Then,

$$\frac{\theta_o(s)}{\theta_i(s)} = P(s),$$
$$\theta_o(s) = P(s)\theta_i(s).$$

To obtain θ_o as a function of time, it is necessary only to transform the function $P(s)\theta_i(s)$ to a time function by means of the inverse Laplace transformation.

3. The transfer function concept provides a convenient way of obtaining the over-all equation of a system from the transfer functions of component units. Let us consider the system represented in Fig. 9·2. Reading from left to right, the output of each unit is the input for the next. The over-all transfer function $E_4(s)/E_1(s)$ may be readily found, since it will be the product of the component transfer functions. Thus,

$$\frac{E_4(s)}{E_1(s)} = P_1(s)P_2(s)P_3(s). \tag{9}$$

4. A special case of the general transfer function $P(s)$ is the frequency transfer function $P(j\omega)$. The independent variable (s) of the transfer function $P(s)$ is complex and may be regarded as equivalent to $\sigma + j\omega$, with σ the real part and $j\omega$ the imaginary part of the complex variable (s). If σ is taken as 0, $s = j\omega$ and can be substituted in the transfer function in place of (s). Accordingly, Eq. (8) becomes

$$\frac{\theta_o(j\omega)}{\theta_i(j\omega)} = \frac{1}{Jj^2\omega^2 + fj\omega} = \frac{1}{-J\omega^2 + fj\omega}. \tag{10}$$

This form of the equation can be designated as $P(j\omega)$. Since ω can be regarded as representing the angular frequency of a sinusoidal function,[1] this form of the transfer function leads to the representation of response of a system as a Fourier spectrum. The means for computing this Fourier spectrum from the transfer function will be described later. Thus $P(s)$ and $P(j\omega)$ may be regarded as two alternative forms of the transfer function, and the discussion given for the $P(s)$ form holds also for the $P(j\omega)$ form.

9·4. Generalized Block Diagrams and Components of Servo Systems. Figure 9·3 shows the chief functional units of a servo system together with the symbols that will be used to represent the different parts of the system and the signals occurring at various points. These symbols may be defined as follows:

[1] The basis for this relationship is given in Table 9·1*b*. It depends essentially on the nature of those transform pairs which show the correspondence of the imaginary part of the roots of the characteristic equation and the time function for which it stands.

$\theta_i(t)$ = the input time function signal, or disturbance,
$\theta_o(t)$ = the output time function signal, or response,
$E(t)$ = the servo error, defined as equal to $\theta_i(t) - \theta_o(t)$,
$H(s)$ = the operator or transfer function representing the controlled member,
$C(s)$ = the controller operator or transfer function. It represents the effect that the controller member exercises on the error.

The physical controller may be defined as including all the units involved in conversion of the error signal to the forcing function that is applied to the output member. In the usual case, in which the output member is a mechanical load, its forcing function (or input) is a torque, and the controller will incude an error-corrective network, amplifier, and

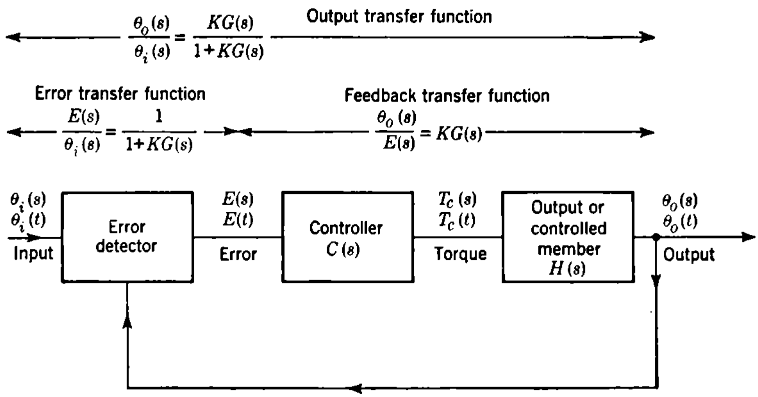


FIG. 9·3.—Block diagram of servo system.

motor. Corrective networks are those which are inserted in a system to improve performance properties through suitable modification of the frequency-response curves of the system and therefore of its transient or steady-state response. They are called error corrective networks if they operate directly on the error signal.

In the simplest servo system, that characterized by a proportional controller, corrective networks of any kind are absent and the motor torque is directly proportional to the error. A derivative controller is one in which the corrective network operates on the error signal to give its (approximate) first or second derivative or both, and an integral controller one in which the corrective network gives the integral of the error signal. Brown designates a servo system as Type 1, 2, or 3, depending on whether it contains a proportional, derivative, or integral controller, respectively. Thus the type of corrective network present is the customary basis on which the controller and the total system are classified and named.

9·5. Interrelations among the Transfer Functions of a Servo System. For mathematical representation of a servo system, three different forms of transfer function have been found useful: the feedback transfer function $\theta_o(s)/E(s)$, the output transfer function $\theta_o(s)/\theta_i(s)$, and the error transfer function $E(s)/\theta_i(s)$ (see Fig. 9·3). The ratios given in each case constitute the definition of the function. It is easily shown that the output and error transfer functions can be written in terms of the feedback transfer function. These relations are given by Eqs. (12) and (14), with $KG(s)$ being used to represent $\theta_o(s)/E(s)$, following the notation of Hall.

To obtain the error function $E(s)/\theta_i(s)$:
let

$$\frac{\theta_o(s)}{E(s)} = KG(s), \tag{11}$$

but

$$E(s) = \theta_i(s) - \theta_o(s), \qquad \text{so} \qquad \theta_o(s) = \theta_i(s) - E(s).$$

Substituting in Eq. (11) for $\theta_o(s)$,

$$\frac{\theta_i(s) - E(s)}{E(s)} = KG(s).$$

Solving for $E(s)/\theta_i(s)$,

$$\frac{E(s)}{\theta_i(s)} = \frac{1}{1 + KG(s)}. \tag{12}$$

To derive the expression for the output function $\theta_o(s)/\theta_i(s)$:

$$\frac{\theta_o(s)}{\theta_i(s)} = \frac{\theta_o(s)}{E(s)}\frac{E(s)}{\theta_i(s)} = KG(s)\frac{1}{1 + KG(s)}, \tag{13}$$

$$\frac{\theta_o(s)}{\theta_i(s)} = \frac{KG(s)}{1 + KG(s)}. \tag{14}$$

9·6. Standard Types of Input Function.—To test a given servo system, either in concrete physical form or in its symbolic counterpart, it is necessary to apply an input disturbance $\theta_i(t)$ which will represent signals or disturbances to which the system will be subjected in actual use. Though the most suitable test functions would seem to be a representive sampling of those expected to occur in practice, this is frequently not possible, owing to the varied nature of such signals. It has, therefore, been customary to use certain standard test signals which provide a basis for estimating how the system will behave under conditions of actual use. Following is a list of the chief standard test signals or conditions, together with their Laplace transforms. The form of the first three time functions is illustrated in Fig. 9·4.

Step Displacement (Heaviside Unit Function):

$$\theta_i(t) = 1, \qquad t > 0$$
$$= 0, \qquad t \leqq 0$$
$$\theta_i(s) = \frac{1}{s}.$$

The time function may be represented by $u(t)$, following the terminology of Gardner and Barnes.[1] In the Heaviside calculus, it is represented by a special symbol meaning unit step function.

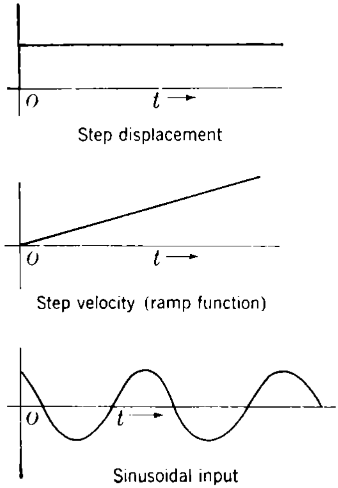


FIG. 9·4.—Standard types of input function $\theta_i(t)$.

Step Velocity Input:

$$\theta_i(t) = Nt, \qquad t > 0,$$

where N is a constant.

$$\theta_i(s) = \frac{N}{s^2}.$$

McColl proposes the term "ramp function" for this input function, following a suggestion of Darrow.[2]

Sinusoidal Input:

$$\theta_i(t) = \cos \omega t \qquad \text{or} \qquad \theta_i(t) = \sin \omega t,$$
$$\theta_i(s) = \frac{s}{s^2 + \omega^2}, \qquad \text{and} \qquad \theta_i(s) = \frac{\omega}{s^2 + \omega^2}.$$

Initial Error:

$$\theta_i(t) = 0; \qquad \theta_o(t) = \phi; \qquad E(t) = \theta_i(t) - \theta_o(t) = -\phi$$

This condition is useful in testing the adequacy of a system as a regulator,[3] in the narrow sense of maintaining the regulated variable at a constant level.

DETERMINATION OF SYSTEM PERFORMANCE

Let us suppose that a provisional selection has been made of the components of a servo system and a block diagram drawn to represent their arrangement. The first question that must be answered is, How will the system perform? What are its response properties? Two

[1] Gardner and Barnes, *Transients in Linear Systems*, Vol. I, Wiley, New York, 1942, pp. 100, 115.

[2] McColl, *Servomechanisms*, Van Nostrand, New York, 1945, p. 38.

[3] For definition, see Sec. 8·3.

alternative approaches may be used. In the one, referred to as the *transient approach*, the response is obtained in terms of the variation of output or error as a function of time for various standard input functions. In the other, designated as the *frequency* or *sinusoidal steady-state approach*, performance is represented in terms of frequency-response curves of output θ_o, relative to error or input, when the system is excited by sinusoidal test functions. Both approaches will be described here in some detail, the transient approach in the present chapter and the frequency approach in the following chapter.

TRANSIENT ANALYSIS

Two principal steps are involved in determination of the response of a given system by means of the transient method of analysis: setting up the differential equation representing the system and solution of this equation to show either the error E or the output signal θ_o as a function of time. The first step, that of setting up the equation of the system, is not limited to the transient method of analysis, since it is also a preliminary to the frequency approach. It is in the second step, in the *solution* of this equation, that one is dealing specifically with the transient method of analysis. It obtains its name from the fact that the complete solution of the differential equation of the system exhibits the characteristics of the transient part of the solution as well as the steady-state part.

9·7. Setting Up the Differential Equation.—The bulk of servo theory is based on the assumption that the system being analyzed is linear, more specifically, that it is one which can be represented by a linear differential equation with constant coefficients.[1] A brief discussion of the limitations involved in this assumption and of the status of nonlinear theory is given in Sec. 11·11. The procedure for setting up the differential equation representing a servo system is therefore no different in nature from that used in writing the equations for any linear system, such as those of electric circuit theory. From the known structure of the system and the physical laws governing the phenomena in different parts, a set of equations is written for different functional units. By appropriate operations, these equations are finally, in the case of a system of one degree of freedom, reduced to a single differential equation containing two functions, beside the independent variable time. In the case of servo systems, these two functions will be the input signal $\theta_i(t)$ and either the error $E(t)$ or the output signal $\theta_o(t)$, depending on which is of more practical interest. The error $E(t)$ will usually be of greater interest, since determination of this function will indicate the extent to which

[1] Our subsequent reference to linear systems will be to linear systems in this narrow sense.

the servo system deviates from an ideal system. In an ideal system $E(t)$ will always be zero.

As a substitute for the procedure summarized above, an alternative procedure involves application of the transfer function concept. In the case of a complex system, it is simpler algebraically and gives one a clearer perspective on the sequence of operations required for obtaining the final differential equation of the system. The procedure consists of the following steps: (1) drawing the block diagram of the system with the various units connected in cascade, (2) determining the transfer function of each of the units represented in the block diagram in the way already described in Sec. 9·3, (3) combining these component transfer functions by multiplication to give the feedback transfer function $\theta_o(s)/E(s)$, and (4) deriving the error transfer function $E(s)/\theta_i(s)$ by means of Eq. (12). If desired, this error equation can easily be rewritten in terms of derivatives to show its identity with the system differential equation derived by the more conventional procedures. The foregoing procedure may be illustrated by means of the following example of a proportional servo system.

Step 1. *Block Diagram of the System.*—For our example the block diagram is given by Fig. 9·3. It shows three units: the error detector or error-measuring means; the controller, comprising here amplifier and motor; and the output member or load.

Step 2. *Determining Transfer Functions of Components.*—First write the equations of the three units. These are given by Eqs. (15) to (17).

$$\theta_i(t) - \theta_o(t) = E(t) \tag{15}$$

for the error measuring means,

$$T_c(t) = k_0 E(t) \tag{16}$$

for the proportional controller, and

$$J\frac{d^2\theta_o}{dt^2} + f\frac{d\theta_o}{dt} = T_c(t) \tag{17}$$

for the controlled member, in this case the load driven by motor torque, where

k_0 = the proportionality constant of controller, in pound-feet per radian error,
J = the inertia of load, slug-ft², in and
f = the viscous friction of the load, in lb-ft per radian per sec.

The remaining symbols have the meanings given in Sec. 9·4.

Equation (16) states that the torque provided by the controller is proportional to the error. For a more complex controller, this equation will be a differential equation or contain an integral term. Equation (17) states that the torque applied to the load is balanced by the opposing torques due to acceleration plus that due to viscous friction. One next transforms each equation of interest to operational form and solves for the ratio of output to input. In the present instance, since we wish initially to obtain the feedback transfer function $\theta_o(s)/E(s)$, we need consider only Eqs. (16) and (17).

From Eq. (16) we thus obtain

$$\frac{T_c(s)}{E(s)} = k_0. \tag{18}$$

From Eq. (17) we obtain, by application of the Laplace transformation and a little algebra,

$$Js^2\theta_o(s) + fs\theta_o(s) = T_c(s), \tag{19}$$

$$\frac{\theta_o(s)}{T_c(s)} = \frac{1}{Js^2 + fs}. \tag{20}$$

Step 3. *Combination of Component Transfer Functions.*—Combining transfer functions given by Eqs. (20) and (18), we obtain

$$\frac{\theta_o(s)}{T_c(s)}\frac{T_c(s)}{E(s)} = \frac{1}{Js^2 + fs}k_0$$

and therefore

$$KG(s) = \frac{\theta_o(s)}{E(s)} = \frac{k_0}{Js^2 + fs}. \tag{21}$$

This is the feedback transfer function.

Step 4. *Derivation of Error Transfer Function.*—Application of Eq. (12) gives

$$\frac{E(s)}{\theta_i(s)} = \frac{1}{1 + KG(s)} = \frac{1}{1 + \dfrac{k_0}{Js^2 + fs}} \tag{22}$$

$$= \frac{Js^2 + fs}{Js^2 + fs + k_0}, \tag{22a}$$

$$E(s) = \frac{Js^2 + fs}{Js^2 + fs + k_0}\theta_i(s) \tag{23}$$

This equation gives, in operational form, the error as a function of the parameters of the system, the complex variable s, and the input function $\theta_i(s)$.

Formulated in terms of derivatives, this equation can be written as

$$J\frac{d^2E(t)}{dt} + f\frac{dE(t)}{dt} + k_0E(t) = J\frac{d^2\theta_i(t)}{dt} + f\frac{d\theta_i(t)}{dt}, \tag{24}$$

where E and θ_i are functions of t. This form is derived from Eq. (23) by multiplying both sides of the equation by the characteristic function, in order to relate the operational expression to the appropriate variable, and then replacing the terms in operational symbols by the appropriate differential expressions. The procedure is merely the reverse of that followed in obtaining the transfer function from the differential equation. Thus, from Eq. (23) we obtain

$$\left.\begin{aligned} (Js^2 + fs + k_0)E(s) &= (Js^2 + fs)\theta_i(s), \\ Js^2E(s) + fsE(s) + k_0E(s) &= Js^2\theta_i(s) + fs\theta_i(s), \\ J\frac{d^2E(t)}{dt} + f\frac{dE(t)}{dt} + k_0E(t) &= J\frac{d^2\theta_i(t)}{dt} + f\frac{d\theta_i(t)}{dt}. \end{aligned}\right\} \tag{25}$$

In most cases, one will not wish to rewrite Eq. (23) in terms of derivatives, since it will itself be the starting point for the subsequent steps. If, however, the problem specifies initial conditions for $E(t)$ and $\theta_i(t)$ and their first derivatives that are different from zero, then it is probably clearest to prepare for the insertion of these initial conditions by first writing the equation in the form of a differential equation in t [Eq. (24)] and then applying the rules for the Laplace transformation of the derivative terms to each term in turn. These rules provide for the insertion of initial conditions (see Table 9·1a).

9·8. Complete Solution of the Differential Equation of the System.—Solution of the system equation, as given by either Eq. (23) or (24), means finding an equation that gives the error E as a function of time t. A plot of this equation provides a graphical picture of the response, which may be compared with a plot of the required performance. Finding such a solution, however, depends on first substituting some function for $\theta_i(t)$ or $\theta_i(s)$, either one of the standard input functions described in Sec. 9·6, or a function representing some arbitrary input of practical interest. The solution therefore will show what the error response is to this particular input function $\theta_i(t)$. The points of special interest in the present section may be conveniently classified under the following topics: (1) the sequence of steps involved in obtaining a solution, (2) relational and nondimensional parameters, (3) nondimensional response curves, and (4) procedures for handling more complex problems. Under the first three topics, discussion will be limited to the simplest possible system—a proportional servo—since the fundamental techniques are involved here as well as in the more complex systems and can thus be made to stand out more clearly. Under the fourth topic, the nature of the supplementary procedures introduced to take care of more complex problems will be considered.

Steps Involved in Obtaining a Solution.—As is well known, a number of alternative procedures are available for finding a solution for differential equations with constant coefficients. Among these are the traditional methods described in any text on differential equations and the operational methods, such as the operational calculus of Heaviside or the related procedures involving the Laplace or Fourier transformations. The present exposition employs the Laplace transformation, as already indicated. Except for differences pointed out in Sec. 9·2, similar steps would be required if one of the other operational methods were employed. Assuming that we start with a differential equation representing the system, three chief steps are required in obtaining a solution: (1) transformation of the differential equation to operational form, (2) algebraic manipulation of the operational equation in order to change it to a form suitable for application of the inverse transformation, and (3) application of the inverse Laplace transformation to yield the required variable as a function of time. These three steps are illustrated below for Eq. (24), representing a proportional servo system.

1. Transformation of differential equation. Application of the Laplace transformation to Eq. (24) yields Eq. (26). Initial conditions [i.e. the value of $E(t)$ and $\theta_i(t)$ and their first derivatives at $t = 0$] are here assumed equal to zero, in order to simplify the algebra. When any of these initial conditions are different from zero, they may easily be inserted in the equation by means of the formal rules for application of the Laplace transformation to derivatives (see Table 9·1*a*). Equation (26) may now be manipulated like any ordinary algebraic equation.

$$Js^2E(s) + fsE(s) + k_0E(s) = Js^2\theta_i(s) + fs\theta_i(s). \tag{26}$$

2. Manipulation of operational equation. Our purpose is next to write Eq. (26) in a form that will permit us to obtain the error E as a function of time when the inverse Laplace transformation is applied. Two steps are therefore necessary: solving Eq. (26) for the variable or function of interest $E(s)$ and recasting the resultant equation in a form to permit the inverse transformation to be evaluated. Let us first solve for the required variable $E(s)$. From Eq. (26) we obtain

$$E(s) = \frac{Js^2 + fs}{Js^2 + fs + k_0}\,\theta_i(s). \tag{27}$$

It will be noted that this equation is the same as Eq. (23), obtained initially by the transfer function procedure of setting up the equation of the system. If we now substitute some specific function for $\theta_i(s)$, we will be in a position to attempt the inverse transformation. Let us assume that the input function $\theta_i(t)$ is a step function. Therefore $\theta_i(s)$ equals $1/s$, and

$$E(s) = \frac{(Js + f)s}{Js^2 + fs + k_0}\,\frac{1}{s} \tag{28}$$

$$= \frac{Js + f}{Js^2 + fs + k_0} = \frac{P(s)}{Q(s)}. \tag{29}$$

To determine an appropriate form for Eq. (29) prior to application of the inverse transformation it is expedient to consider the fundamental forms that appear under $F(s)$ in a table of transform pairs. Inspection of Table 9·1*b* shows that the form of the *time function* corresponding to a given function of s, such as $E(s)$ of Eq. (29), depends on the poles of the function. These are determined, it will be recalled, by determining the roots of the equation $Q(s) = 0$. This equation is known as the *characteristic equation*, and the polynomial $Q(s)$ as the characteristic function.[1] The roots may be more clearly displayed by writing $Q(s)$ as a product of factors containing the roots.

$$E(s) = \frac{P(s)}{(s + s_1)(s + s_2)}. \tag{30}$$

As a final step Eq. (30) may be written as a sum of partial fractions.

$$\frac{P(s)}{(s + s_1)(s + s_2)} = \frac{A}{s + s_1} + \frac{B}{s + s_2}, \tag{31}$$

where A and B are constants and $-s_1$ and $-s_2$ are the roots of $Q(s) = 0$. Several methods are available for evaluating the constants, or coefficients, A and B. One basic method, the method of undetermined coefficients, will be found described in any standard algebra.

A second method, corresponding to a procedure commonly used for evaluating the residues at the poles of a complex rational function, is almost apparent by inspection of Eq. (31). To determine A, we may multiply both sides of the equation by $(s + s_1)$. This new equation will be true for all values of s, hence for $s = -s_1$, which may be substituted for s throughout.

$$\left.\frac{P(s)}{(s + s_2)}\right]_{s=-s_1} = A + \frac{B(-s_1 + s_1)}{-s_1 + s_2} \tag{32}$$
$$= A + 0.$$

Thus terms on the right other than A are eliminated and the value of A can be found from the left-hand side of the equation. A like procedure can be used for evaluating the other constant B. Analogous though slightly modified methods are applicable when any one of the denominators on the right-hand side of the equation contains higher degree terms in s, such as $s^2 + \omega^2$.[2] The same procedure may be followed regardless of the number of factors, and therefore constants, involved.

Since in our example the characteristic function is of second degree, the roots obtained by setting it equal to zero are readily found.

$$Js^2 + fs + k_0 = 0;$$
$$s = -\frac{f}{2J} \pm \sqrt{\left(\frac{f}{2J}\right)^2 - \frac{k_0}{J}}. \tag{33}$$

It will be found convenient later if we now write the above roots as complex numbers, by factoring out $j = \sqrt{-1}$ from the radical.

$$s = -\frac{f}{2J} \pm j\sqrt{\frac{k_0}{J} - \left(\frac{f}{2J}\right)^2}. \tag{34}$$

[1] See Gardner and Barnes, *op. cit.*, p. 132.

[2] For a detailed description of procedures, see Gardner and Barnes, *op. cit.*, pp. 153–164.

3. Application of inverse Laplace transformation. We are now in a position to determine the different types of time response that may occur in the case of an error function such as Eq. (29) with a second-degree characteristic equation. Consideration of Eq. (34), giving the roots of a quadratic, together with our table of transform pairs will show what kinds of time response are possible and the reason why any given one occurs. We shall find that three types of time response may occur: (*a*) an undamped sinusoidal oscillation, (*b*) a damped oscillation, and (*c*) a nonoscillatory damped exponential. (These responses are illustrated in Fig. 9·5 for both the error and the output response.) The type

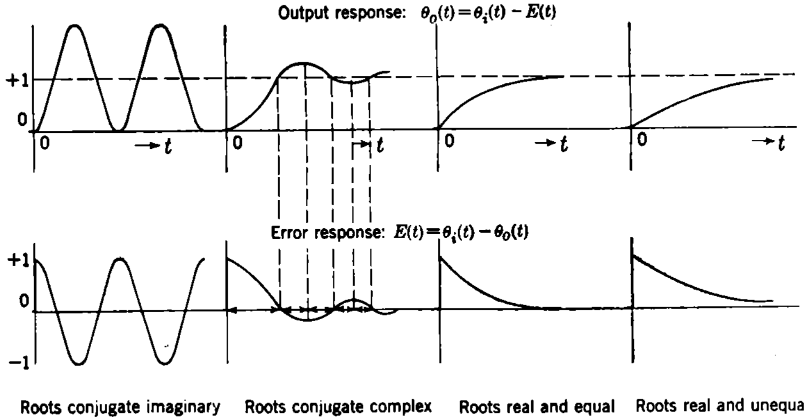


FIG. 9·5.—Types of response curves corresponding to different solutions of a second-degree characteristic equation.

of response occurring will depend on the character of the roots of our quadratic. These in turn will depend on the value of $f/2J$ relative to $\sqrt{k_0/J}$ [see Eq. (34)]. When f and therefore $f/2J = 0$, the two roots of the quadratic are conjugate imaginaries and the time response is an undamped oscillation. When $f/2J$ is greater than zero but less than $\sqrt{k_0/J}$, the roots are conjugate complex and the time response is a damped oscillation. When $f/2J = \sqrt{k_0/J}$, the roots are real and equal and the error time response is a decreasing exponential. Finally, when $f/2J > \sqrt{k_0/J}$, the two roots are real and unequal and the time response is the sum of two exponentials. These results are illustrated in Fig. 9·5. The curve corresponding to $f/2J = \sqrt{k_0/J}$ is designated as critically damped, since it represents the smallest value of $f/2J$ relative to $\sqrt{k_0/J}$ at which oscillations will not occur. Curves corresponding to still greater values of $f/2J$ are designated as overdamped. These results may be readily obtained by substituting different values of $f/2J$ into Eq. (34) and determining the nature of the roots resulting. Then by examination of our table of transform pairs the nature of the time response corresponding to different types of roots may be determined. For purposes of illustration this procedure is carried out in two of the special cases.

1. Undamped Oscillations. When $f/2J = 0$, $s = \pm j\sqrt{k_0/J}$. There are thus two conjugate imaginary roots, representing two conjugate poles on the imaginary axis. Transform pair (*c*) of Table 9·1*b* shows that the corresponding time function is an undamped oscillation of frequency $\sqrt{k_0/J}$. This frequency is called the undamped natural frequency and represented by the symbol ω_n.

If we wish to evaluate formally the inverse transform of $E(s)$, we find, substituting $f = 0$ in Eq. (29),

$$E(s) = \frac{Js}{Js^2 \pm k_0} = \frac{s}{s^2 + \dfrac{k_0}{J}} = \frac{s}{s^2 + \omega_n^2}. \tag{35}$$

Then upon applying transform pair (*c*), there results the inverse Laplace transform

$$E(t) = \cos \omega_n t. \tag{36}$$

This equation gives the complete solution of our differential equation under the condition specified.

2. Damped oscillations. We again start with the error equation for a step-function input, given by Eq. (29) and repeated here:

$$E(s) = \frac{Js + f}{Js^2 + fs + k_0} = \frac{s + \dfrac{f}{J}}{s^2 + \dfrac{f}{J}s + \dfrac{k_0}{J}}. \tag{29}$$

The roots of the characteristic equation are

$$\begin{aligned} s &= -\frac{f}{2J} \pm j\sqrt{\frac{k_0}{J} - \left(\frac{f}{2J}\right)^2} \\ &= -\alpha \pm j\omega \end{aligned}$$

Then Eq. (29) can be written

$$E(s) = \frac{s + 2\alpha}{(s + \alpha - j\omega)(s + \alpha + j\omega)} = \frac{s + 2\alpha}{(s + \alpha)^2 + \omega^2}.$$

Applying transform pair 1.303, page 342, Gardner and Barnes,

$$\left.\begin{aligned} E(t) &= \left(1 + \frac{\alpha^2}{\omega^2}\right)^{1/2} e^{-\alpha t} \sin(\omega t + \phi) \\ \phi &= \tan^{-1}\frac{\omega}{\alpha}. \end{aligned}\right\} \tag{37}$$

Relational and Nondimensional Parameters.—The outline given of the steps involved in solving the system equation was limited for reasons of concreteness to the dimensional parameters characterizing specific physical components. Examples of such parameters are the moment of inertia J, measured in slug ft^2 (or gm-cm^2), the viscous friction f, measured in lb-ft (or dyne-cm) per radian per sec and the stiffness or proportionality constant k_0, measured in lb-ft (or dyne-cm) per radian error. It has been found of value in many problems to employ instead what might be called *relational parameters*, i.e., parameters that depend on relations or ratios among the dimensional parameters rather than on any one physical property. In the present section a few of the more important of these parameters will be considered and an attempt made to show how they arise naturally out of the different types of solution considered above. In some instances these parameters are nondimensional, a feature that contributes greatly to their value.[1] In other instances, though not in themselves nondimensional, they provide a basis for easily forming new nondimensional parameters through combination with other parameters or variables having the inverse dimensions.

Let us consider, first, a pair of parameters that are defined in relation to the amount of damping in a system, namely, the *damping ratio* ζ and the *undamped natural angular frequency* ω_n. It was found possible in the previous section to arrange the different types of time response in a series corresponding to the magnitude of $f/2J$ relative to $\sqrt{k_0/J}$. As $f/2J$ increases from zero to values greater than $\sqrt{k_0/J}$, the time response was found to change progressively from an oscillation of constant amplitude to damped oscillations that are damped out more and more rapidly and finally to nonoscillatory responses compounded of exponentials. If, now, we wish to indicate the position in such a series at which the response of a given system will lie, it seems natural to select as a reference point one that will separate the oscillatory from the nonoscillatory responses. Such a reference point is provided by the response that occurs when the roots of the characteristic equation are real and equal. It is known as the *critically damped response* and will occur when

$$\frac{f}{2J} = \sqrt{\frac{k_0}{J}}.$$

For smaller values of $f/2J$ oscillatory responses occur; for greater values, nonoscillatory responses. We may refer to f as the damping parameter of the system, since with all other parameters constant, variation of it alone will provide the series of responses mentioned above. Similarly, f_c may be used to designate the value of f for critical damping. Then ζ

[1] A clear discussion of nondimensional equations may be found in D. P. Campbell, "Theory of Automatic Control Systems," *Industrial Aviation*, September 1945, p. 62-

is designated as the damping ratio and is defined as the ratio f/f_c, or the ratio of actual damping to critical damping. It is convenient to be able to express ζ in terms of the whole set of system parameters. As a preliminary step, f_c is written in terms of system parameters. When $f = f_c$, i.e., for critical damping,

$$\left.\begin{aligned} \frac{f_c}{2J} &= \sqrt{\frac{k_0}{J}}; \\ \text{therefore} \qquad f_c &= 2\sqrt{k_0 J} \\ \text{and} \qquad \zeta &= \frac{f}{f_c} = \frac{f}{2\sqrt{k_0 J}}. \end{aligned}\right\} \tag{38}$$

The quantity ζ now provides a convenient index of the position of a given response in the series of possible responses. For a critically damped system, $\zeta = 1$, as is apparent from Eq. (38). For $\zeta < 1$, oscillatory or underdamped responses occur; for $\zeta > 1$, nonoscillatory or overdamped responses occur.

The related parameter ω_n, the undamped angular frequency, has already been defined as equal to $\sqrt{k_0/J}$. It is the value of the frequency ω when f and therefore ζ equal zero.

These new parameters ζ and ω_n may now be substituted for the "physical parameters" J, f, and k_0 appearing in any of the equations occurring in our analysis of the proportional servomechanism. Thus in place of Eq. (27), the error equation can be written as

$$E(s) = \frac{s^2 + 2\zeta\omega_n s}{s^2 + 2\zeta\omega_n s + \omega_n^2}\,\theta_i(s). \tag{39}$$

From this point, we might proceed exactly as before and find the various possible time solutions in terms of ζ and ω_n for any given input function. This is in fact the more usual procedure in expositions already published.[1] If, however, the solutions have already been obtained in terms of J, f, and k_0, it may be simpler to use them as the point at which the substitutions are made.[2] The equations for the different types of response to step functions may be written as in the following equations:[3]

[1] G. S. Brown, "Transient Behavior and Design of Servomechanisms," NDRC Sec. D-2 Report, November 1940; H. Harris, "The Analysis and Design of Servomechanisms," NDRC Report, 1942; C. S. Draper and G. V. Schliestett, "General Principles of Instrument Analysis," *Instruments*, **12,** 137–142, 1939.

[2] In making the substitutions in any equation, it will be found convenient to perform such algebraic operations as will lead to coefficients such as f/J and k_0/J and to substitute for these ratios terms containing ω_n^2 and ζ. Thus, it is readily shown from the definitions of ω_n and ζ that $k_0/J = \omega_n^2$ and $f/J = 2\zeta\omega_n$.

[3] *Cf.* Brown, *op. cit.*, Table IV.

$$E(t) = \theta_i \cos \omega_n t \qquad \text{for } \zeta = 0, \tag{40}$$

$$E(t) = \theta_i \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} \sin\left(\sqrt{1-\zeta^2}\,\omega_n t + \phi\right) \qquad \text{for } 0 \leqq \zeta < 1, \tag{41}$$

$$= \theta_i(1 + \omega_n t)e^{-\omega_n t} \qquad \text{for } \zeta = 1, \tag{42}$$

$$= \theta_i e^{-\zeta\omega_n t}\left(\cosh\sqrt{\zeta^2-1}\,\omega_n t + \frac{\zeta}{\sqrt{\zeta^2-1}}\sinh\sqrt{\zeta^2-1}\,\omega_n t\right) \qquad \text{for } \zeta > 1, \tag{43}$$

where θ_i is the magnitude of step function $\theta_i(t)$ and $\phi = \tan^{-1}\sqrt{1-\zeta^2}/\zeta$.

It will be shown in the following subsection how equations such as these provide a basis for construction of sets of generalized response curves that may be used to simplify the determination of system response.

A second pair of relational parameters frequently useful is k_v, the velocity error constant of the system, and T or τ, the so-called time constant or characteristic time of some portion of the system. T is frequently written with a subscript to indicate the portion of the system represented when such time constants appear in various parts of the system. Thus T_m stands for the motor time constant. (The motor time constant is the time required for the motor to reach 63.3 per cent of its final speed after the application of a step input voltage.) In the case of the proportional servomechanism, k_v is defined as k_0/f and has the dimension of 1/time, and T is defined as J/f and has the dimension of time.[1] One way of indicating the origin of these parameters is to start with the feedback transfer function as given by Eq. (21) and divide the numerator and denominator by f. Substitution of k_v and T as defined above will now permit elimination of the three original physical parameters.

$$\frac{\theta_o(s)}{E(s)} = \frac{k_0}{Js^2 + fs} = \frac{\dfrac{k_0}{f}}{\dfrac{J}{f}s^2 + s} = \frac{k_v}{Ts^2 + s}. \tag{44}$$

If this transfer function is used in determination of types of error response, the response equations will, of course, be given in terms of k_v and T instead of k_0, J and f, or ω_n and ζ. The parameters to be preferred in a given problem will naturally depend on the nature of the application. Thus if one is interested in the magnitude of the velocity lag error it is convenient to use the velocity error constant k_v as a parameter, since $E_v = N/k_v$, where E_v is the velocity lag error and N is the slope of the input ramp function.[2]

[1] The symbol T for time constant should not be confused with $T_c(t)$, used for torque output of controller.

[2] For derivation of this equation see Sec. 9·9.

Nondimensional Response Curves.—The solution of a system equation in terms of relational parameters which can readily be converted into nondimensional parameters leads directly to the construction of nondimensional graphs of the various possible forms of response. This approach has been developed in this country by Draper and his students in the analysis of instruments and extended to the analysis of servomechanisms by Brown and his students and others. The procedure is of great value in that the same set of response curves may be used to represent systems that, although equivalent in organization and hence in the form of the system equations, may differ widely in the values of the physical parameters. For a given type of system, one may refer to the appropriate set of curves and determine the form of the response and its duration and magnitude in terms of nondimensional variables. The nondimensional magnitudes and durations may then be converted into dimensional terms, and the actual characteristics of the response determined. The curves may also be used in the inverse problem of determining the design parameters required for a system to meet given performance specifications. For a detailed account of this approach, the papers by Draper, Brown, and their co-workers may be consulted.[1]

The construction and use of nondimensional response curves may be illustrated here for the proportional servo system discussed above. The equations for the error response to a step function are given by Eqs. (40) through (43), with E the dependent variable, t the independent variable, and ζ and ω_n the parameters. If both sides of the equations are divided by θ_i, the magnitude of the input step function, and $\omega_n t$ is taken as the independent variable, then the equations will be nondimensional. By way of illustration, Eq. (45) shows the result of carrying out these operations on Eq. (41).

$$\frac{E(t)}{\theta_i} = \frac{e^{-\zeta(\omega_n t)}}{\sqrt{1-\zeta^2}} \sin\left[\sqrt{1-\zeta^2}\,(\omega_n t) + \phi\right] \tag{45}$$

for $0 \leqq \zeta < 1$.

[1] C. S. Draper and G. V. Schliestett, "General Principles of Instrument Analysis," *Instruments*, **12**, 137–142 (1939); C. S. Draper and G. P. Bentley, "Design Factors Controlling the Dynamic Performance of Instruments," *Trans. ASME*, **62**, No. 5, 421–432, July 1940; G. S. Brown, *op. cit.*, pp. 1–47. References to the German literature dealing with the use of dimensionless ratios may be found in E. S. Smith, *Automatic Control Engineering*, McGraw-Hill, 1943, pp. 343*ff.* Material on the theoretical background in the field of dimensional and nondimensional analysis may be found in P. W. Bridgman, *Dimensional Analysis*, Yale University Press, New Haven, 1931. Recent expository accounts that may also prove useful are those of H. Sohon, *Engineering Mathematics*, Chap. 3, "Dimensional Analysis," Van Nostrand, New York, 1944; and E. R. Van Driest, "On Dimensional Analysis and the Presentation of Data in Fluid-flow Problems," Convention ASME, *Paper* 45, Nov. 26, 1945.

The independent variable $\omega_n t$ is dimensionless, for ω_n has the dimensions of 1/sec and t the dimensions of sec. $E(t)/\theta_i$ is dimensionless, since it is the ratio of two quantities with the same dimensions. Equation (45) may now be plotted by the procedure used for plotting any equation.

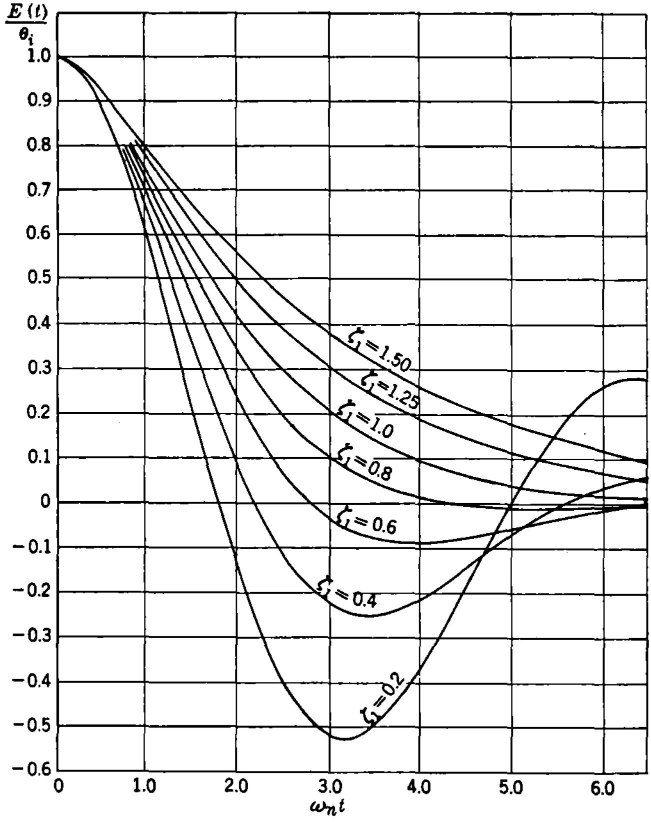


FIG. 9·6.—Dimensionless transient error curves of a servomechanism with a Type I controller when subjected to a suddenly applied input angle θ. (*Courtesy of G. S. Brown, Transient Behavior and Design of Servomechanisms, Massachusetts Institute of Technology,* 1941 *and* 1943.)

For any particular value of the remaining parameter ζ, $\omega_n t$ is given each of a series of numerical values, 0, 1, 2 etc., and the corresponding values of $E(t)/\theta_i$ calculated. The procedure is then repeated for another value of ζ in the allowable range, and so on. The resulting plots of Eqs. (41) to (43) for different values of ζ are given in Fig. 9·6.

Equation (40) represents a limiting case of Eq. (41), which occurs when $\zeta = 0$, and is an oscillation of constant amplitude or zero decrement.

The way that such nondimensional curves may be used to determine the response of a system whose physical parameters are known may be illustrated by the following simple example. Suppose in a proportional servomechanism that $J = 2$ slug-ft², $f = 12$ lb-ft/radian-per-sec and $k_0 = 50$ lb-ft/radian and that the magnitude of the input step function is 10°. Suppose one wishes to plot the error response curve as a function of time and to determine the interval that will elapse before the error is less than 0.4°. The first step is to determine the magnitude of the relational parameters.

$$\zeta = \frac{f}{2\sqrt{Jk_0}} = \frac{12}{2\sqrt{2(50)}} = 0.6;$$

$$\omega_n = \sqrt{\frac{k_0}{J}} = \sqrt{\frac{50}{2}} = 5 \qquad \text{radians per sec.}$$

Turning to Fig. 9·6, it is apparent that the *form* of the response curve is given by the curve for $\zeta = 0.6$. To read durations in seconds, we need to find t from $\omega_n t$, as indicated by Eq. (46).

$$t = \frac{\omega_n t}{\omega_n} = \frac{\omega_n t}{5} \qquad \text{sec.} \tag{46}$$

Thus the nondimensional abscissa values of Fig. 9·6 need only be divided by 5 to give us the time coordinate in seconds. Similarly, the ordinate values can be multiplied by $\theta_i = 10°$ to give the error in degrees. From this scale, the time corresponding to an error of 0.4° or less can be read off. Alternatively, the error specification can be formulated in nondimensional terms as a required error of less than $E/\theta_i = 0.4°/10°$ or 4 per cent. The nondimensional curve shows that the corresponding value of $\omega_n t$ is 5.5. In time units this will correspond to 5.5/5, or 1.1 sec.

In the above example, the relevant nondimensional curves (Fig. 9·6)[1] were those showing the error response resulting from an input *step* function. Figure 9·7[2] shows the family of nondimensional response curves resulting from an input *ramp* function. In this curve the ordinate is given in terms of $E(t)/E_{ss}$ where E_{ss} signifies the steady-state error, which will here be a velocity lag. The abscissa is again given in terms of the dimensionless product, $\omega_n t$. Dimensionless curves for more complex systems, such as the Type 2 (derivative controller) and Type 3 (integral controller) may be found in Brown's report.

Treatment of More Complex Problems.—In the sections above, an outline has been given of the formal procedures involved in solving the differential equation of a servo system. An extremely simple system operating under the influence of a simple set of external agencies was

[1] From Brown, *op. cit.*, p. 44.

[2] *Ibid.*, p. 13.

chosen for purposes of illustration. It will be of interest now to determine how the basic scheme outlined is modified as the external conditions and the system become somewhat more complex. Possible changes in the external agencies consist in (1) variation in the form of the input forcing function and (2) application of accessory disturbances at other points of the system. Possible changes in the system itself consist in (1) changes in the nature of the controller or corrective network and (2) an increase in the number of energy-storage units in the output member or other parts of the system. It will be found that the same

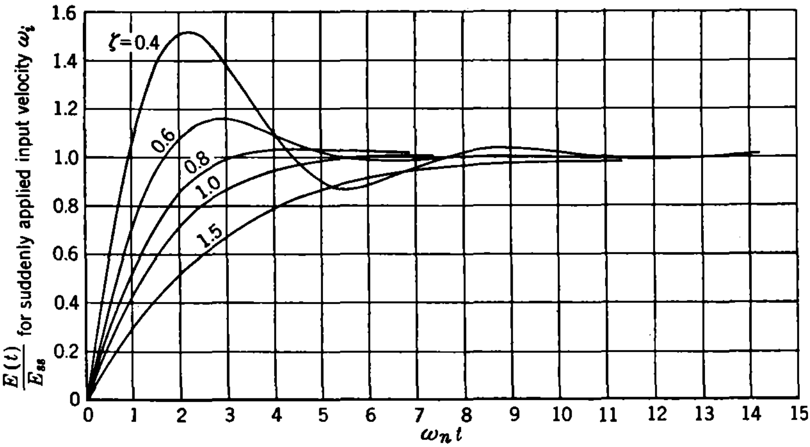


FIG. 9·7.—Dimensionless transient error curves for a servomechanism with a Type 1 controller when subjected to a suddenly applied input velocity. (*Courtesy of G. S. Brown.*)

formal procedure already described for determination of system response is applicable to these more complex problems. Complicating factors are incorporated in the treatment by properly representing the given condition in the differential or operational equations initially set up. After one has solved for the function of interest, such as $E(s)$, it will be found that the expression on the right-hand side of the equation will merely be somewhat more elaborate than in our previous example and may contain additional functions representing external disturbances whose form must be known. But one may proceed as before to recast the equation in a way to facilitate evaluation of the inverse Laplace transformation and finally carry out the transformation. Let us now consider more specifically how the changes just mentioned will be represented in the equations.

Procedures for dealing with *different types of input function* can be disposed of quite easily. A convenient starting point for the discussion is the error transfer function given in general form by Eq. (47) and for the proportional servo system by Eq. (22*a*). Since $KG(s)$ is invariably a rational function, it may be written as $P(s)/Q(s)$. Then

$$\frac{E(s)}{\theta_i(s)} = \frac{1}{1 + KG(s)} = \frac{1}{1 + \frac{P(s)}{Q(s)}} = \frac{Q(s)}{P(s) + Q(s)}, \tag{47}$$

and

$$E(s) = \frac{Q(s)}{P(s) + Q(s)}\,\theta_i(s). \tag{48}$$

In our previous example $\theta_i(s)$ was equal to $1/s$, which was substituted for $\theta_i(s)$. Whatever form $\theta_i(s)$ takes, it may be substituted at this point for $\theta_i(s)$, and any factors in its denominator that are identical with factors of $Q(s)$ may be cancelled out. The resultant expression is then treated in the manner already described until the inverse transform of $E(s)$ has been obtained. Where initial conditions other than zero must be inserted,

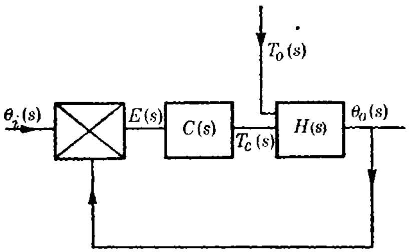


FIG. 9·8.—Block diagram of a servo system involving an output disturbance $T_0(s)$.

the procedure previously discussed (page 239) is applicable regardless of the nature of the input function.

By an *accessory disturbance* is meant any forcing function applied to the system at points other than at the input that may tend to introduce a servo error or interfere with its elimination. Thus in Fig. 9·8, a torque $T_o(t)$ acting on the output member will tend to alter θ_o and hence produce an error. Such disturbances may be handled by incorporating them in the original differential or operational equation set up to represent the system. The final error function will then contain $T_o(s)$ or an equivalent symbol in addition to $\theta_i(s)$. Hence, a preliminary to the manipulation of $E(s)$ and evaluation of its inverse transform will be the substitution of a specific function for $T_o(s)$ as well as for $\theta_i(s)$.

By way of illustration, let us derive a general expression for the error in a servo system acted on by a disturbing torque in addition to $\theta_i(t)$. The block diagram is given by Fig. 9·8, with the various functions written as functions of s.

The transfer function of the controller member is represented by $C(s)$ and of the output or controlled member by $H(s)$. For obtaining

$\theta_o(s)$ in terms of $E(s)$, the equations of interest are those characterizing units between $E(s)$ and $\theta_o(s)$.

$$\left.\begin{aligned} T_c(s) &= C(s)E(s) \\ \theta_o(s) &= H(s)T_a(s) \end{aligned}\right\} \tag{49}$$

where $T_a(s)$ is the Laplace transform of $T_a(t)$ and $T_a(t)$ represents the total torque applied to the output member. Since $T_a(t)$ will equal the sum of the controller torque $T_c(t)$ and the disturbing torque[1] $T_o(t)$,

$$T_a(s) = T_c(s) + T_o(s).$$

Therefore,

$$\left.\begin{aligned} \theta_o(s) &= H(s)[T_c(s) + T_o(s)] \\ &= H(s)[C(s)E(s) + T_o(s)], \\ \theta_o(s) &= C(s)H(s)\,E(s) + H(s)T_o(s). \end{aligned}\right\} \tag{50}$$

This equation gives the output response in terms of the error function $E(s)$ and the torque disturbance $T_o(s)$, in addition to system transfer functions. To obtain an equation comparable to Eq. (23) of our previous example we want to solve for $E(s)$ and to eliminate $\theta_o(s)$. This may be done by substituting $\theta_i(s) - E(s)$ for $\theta_o(s)$. The validity of the substitution follows from Eq. (51), which may be assumed for the error detector.

$$\left.\begin{aligned} E(t) &= \theta_i(t) - \theta_o(t); \\ \text{therefore} \qquad & \\ \theta_o(s) &= \theta_i(s) - E(s). \end{aligned}\right\} \tag{51}$$

Upon making the substitution in Eq. (50) and solving for $E(s)$, we obtain

$$\begin{aligned} E(s) &= \frac{1}{1 + C(s)H(s)}\,\theta_i(s) - \frac{1}{1 + C(s)H(s)}\,H(s)T_o(s) \qquad (52) \\ &= \frac{1}{1 + C(s)H(s)}\,[\theta_i(s) - H(s)T_o(s)]. \end{aligned}$$

This equation, written in terms of operators, is designated by Brown and Hall[2] as the basic equation for a closed-cycle system such as shown in Fig. 9·8. The equation shows the error as a function of the transfer functions of the different units and the excitation functions $\theta_i(s)$ and $T_o(s)$. If the system transfer functions $C(s)$ and $H(s)$ are known and in addition the form of the excitation functions $\theta_i(s)$ and $T_o(s)$, then our equation will contain only known parameters in addition to terms in s.

[1] Some authors take the sign preceding $T_o(s)$ as negative and thus obtain a corresponding reversal of sign in the final equation.

[2] See Brown and Hall, *op. cit.*, p. 3. The symbols used here closely follow those of Brown and Hall. The chief difference lies in the fact that these authors treat p as an operator and thus write equations in which transfer functions appear as functions of p and excitation functions as functions of t.

It may therefore be recast in partial fraction form, and the inverse Laplace transformation applied. The first term of Eq. (52) shows in operational form the part of the error response due to the input function $\theta_i(s)$, and the second term the part due to the accessory disturbance $T_o(s)$. If $T_o(s)$ is assumed to be zero, the second term vanishes and we are left with the basic error equation of a servo system not subject to accessory disturbances.

Finally, in order that we may see more clearly the relation of Eq. (52) to that previously obtained for a proportional servo system, Eq. (23), let us substitute for $C(s)$ and $H(s)$ the transfer functions used in our previous example. There, $C(s) = k_0$ and $H(s) = 1/(Js^2 + fs)$. Therefore,

$$\begin{aligned} E(s) &= \frac{1}{1 + \dfrac{k_o}{Js^2 + fs}}\,\theta_i(s) - \frac{1}{1 + \dfrac{k_o}{Js^2 + fs}}\,\frac{T_o(s)}{Js^2 + fs} \\ &= \frac{Js^2 + fs}{Js^2 + fs + k_0}\,\theta_i(s) - \frac{1}{Js^2 + fs + k_0}\,T_o(s). \end{aligned} \tag{53}$$

If, as before, $\theta_i(t)$ is a step function, then

$$E(s) = \frac{(Js + f)}{Js^2 + fs + k_0} - \frac{1}{Js^2 + fs + k_0}\,T_o(s). \tag{54}$$

Comparison of Eq. (53) with Eq. (23) shows them to be identical except for the second term in Eq. (53), due to the disturbing torque $T_o(s)$. When a known function is substituted for $T_o(s)$, determination of $E(t)$ can proceed as described.

So much for the effect of more complex excitation conditions on our problem. Let us now consider the chief ways in which the system itself may increase in complexity. Two types of possible variation have been mentioned: variations in the type of controller and variations in the type and number of energy storage units. It should be apparent that changes of either type will be reflected in the nature of the transfer functions of various units of the system and hence in the over-all transfer function.

Let us consider first the more common variations that may occur in the *controller*. In the sections dealing with correction of servo-system performance in Chap. 11, development of particular types of controller will be reviewed. Here it will suffice to consider the nature of the transfer functions of two important controller types, the integral and derivative error controllers, and to indicate how their incorporation in a system modifies the error equation. Finally, at the end of this section, the nature and effect of energy storage elements and procedures for solution of the higher-order equations that occur will be reviewed. In Sec. 11·8, we shall find that the *integral type of controller* has been developed as a means for compensating for the steady-state errors of a servo system. The transfer function of such a controller is given by Eq. (55).

$$C(s) = \frac{T_c(s)}{E(s)} = k_0 + \frac{n}{s} = k_0\left[1 + \frac{n}{k_0 s}\right] = \frac{k_0 s + n}{s}, \tag{55}$$

where k_0 and n are constants that indicate the amount of proportional and integral control, respectively. To determine the effect of an integral controller on the error equation of the system, let us start with the basic error equation, as given by Eq. (52). For purposes of simplification, let it be assumed that the disturbing torque $T_o(s) = 0$ and that

$$H(s) = \frac{1}{Js^2 + fs}$$

as in our earlier illustrations. Then

$$E(s) = \frac{Js^2 + fs}{Js^2 + fs + C(s)}\,\theta_i(s). \tag{56}$$

Substituting for $C(s)$ as given by Eq. (55), there results

$$\begin{aligned} E(s) &= \frac{Js^2 + fs}{Js^2 + fs + \dfrac{k_0 s + n}{s}}\,\theta_i(s) \\ &= \frac{(Js + f)s^2}{Js^3 + fs^2 + k_0 s + n}\,\theta_i(s). \end{aligned} \tag{57}$$

The incorporation of an integral controller in the system raises the degree of the characteristic equation by 1. The error transfer function is now of the third order rather than of the second order, as in the case of a proportional or Type 1 servo system.

Another important type of controller is the *derivative error controller*, to be considered in more detail in Sec. 11·4. It is important as a device either for increasing the stability of a system or for reduction of its transient error. The transfer function of an ideal derivative controller, involving proportional, first derivative, and second derivative control is given by

$$C(s) = \frac{T_c(s)}{E(s)} = k_0 + k_1 s + k_2 s^2. \tag{58}$$

Its effect on the error equation of a system is indicated by substituting this function for $C(s)$ in Eq. (56).

$$\begin{aligned} E(s) &= \frac{Js^2 + fs}{Js^2 + fs + (k_0 + k_1 s + k_2 s^2)}\,\theta_i(s) \\ &= \frac{Js^2 + fs}{(J + k_2)s^2 + (f + k_1)s + k_0}\,\theta_i(s). \end{aligned} \tag{59}$$

The incorporation of this type of controller does not alter the degree of the characteristic equation at all but merely changes the value of the coefficients. Solution of the error equation may therefore be carried out by the same procedures as already described.

Let us consider now the effect of components that are commonly referred to as *energy-storage units*, or units responsible for response lags in the system. The effect of such units is to add factors of the form $1/(As + B)$ or $1/s(As + B)$ to the feedback transfer function.

An example of an energy storage unit is the field control circuit of a d-c motor, in which inductance and resistance in series may sometimes be encountered. Let us suppose that the instantaneous motor torque[1] is proportional to the instantaneous field current, the proportionality constant being represented by k_m. The equations of the motor can then be written

$$\left.\begin{aligned} T_c(t) &= k_m i_f(t), \\ e(t) &= L\frac{di_f(t)}{dt} + Ri_f(t), \end{aligned}\right\} \tag{60}$$

where $e(t)$ is the applied voltage, $i_f(t)$ the field current, and $T_c(t)$ the resultant torque. Written as transfer functions, Eq. (60) becomes

$$\begin{aligned} \frac{I_f(s)}{e(s)} &= \frac{1}{Ls + R}, \\ \frac{T_c(s)}{I_f(s)} &= k_m, \end{aligned} \tag{61}$$

so

$$\frac{T_c(s)}{e(s)} = \frac{k_m}{Ls + R}. \tag{62}$$

Another example is provided by the smoothing network of Fig. 9·9a. Its differential equation is

$$Ri(t) + \frac{1}{C}\int i(t)\,dt = e_i(t),$$

so

$$\left(R + \frac{1}{Cs}\right) I(s) = e_i(s),$$

where $e_i(t)$ is the applied voltage and $i(t)$ the current.

$$\frac{I(s)}{e_i(s)} = \frac{1}{R + \dfrac{1}{Cs}} = \frac{Cs}{RCs + 1}.$$

[1] It should be borne in mind that the instantaneous torque referred to here can be measured directly at motor output only when the motor velocity has been brought to zero by an opposing torque. At motor speeds greater than zero the measured torque will decrease as speed increases. This decrease in torque, as shown in the familiar torque-speed curves, can, however, be regarded as due to a counter torque associated with viscous friction. For discussion of this problem, see Sec. 13·2.

But

$$e_0(t) = \frac{1}{C}\int i\,dt,$$

therefore

$$\frac{e_0(s)}{I(s)} = \frac{1}{Cs}.$$

Then,

$$\frac{e_o(s)}{e_i(s)} = \frac{e_o(s)}{I(s)}\frac{I(s)}{e_i(s)} = \frac{Cs}{RCs + 1}\frac{1}{Cs} = \frac{1}{RCs + 1}. \tag{63}$$

The transfer functions of physical networks incorporated as part of derivative or integral controllers are likely to contain factors indicative of such energy storage elements in addition to the function that one desires to synthesize.

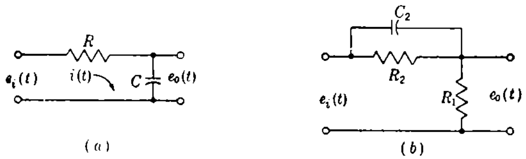


FIG. 9·9.—(a) "Smoothing," or low-pass network; (b) "phase-lead" network.

Thus, one common type of network used as part of a derivative controller (intended to provide proportional and first derivative terms) is that shown in Fig. 9·9b. Its transfer function is given by

$$\frac{e_2(s)}{e_1(s)} = \frac{R_1}{R_1 + R_2}\,\frac{1 + R_2C_2s}{1 + \dfrac{R_1R_2C_2}{R_1 + R_2}s}. \tag{64}$$

It is thus of the form

$$\frac{e_2(s)}{e_1(s)} = k_0\frac{1 + T_2s}{1 + T_1s}, \tag{65}$$

where k_0, T_1, T_2 are constants.

The transfer function for an *ideal* derivative controller is, however, of the form

$$C(s) = k_0(1 + Ts).$$

The factor $1/(1 + T_1s)$ in Eq. (65) thus represents the extent to which a passive network fails to meet the requirements of an ideal derivative controller.

Finally, it will be recalled that the mechanical load of a servo system involving inertia and viscous friction has a transfer function of this same form and is written

$$\frac{\theta_o(s)}{T_c(s)} = \frac{1}{s(Js + f)}.$$

As these examples indicate, factors of the type in question may occur at various points throughout the servo system. The effect, when several such factors are present, will be to increase the degree of the characteristic equation of the feedback and error transfer functions. An example is provided by the system represented in the block diagram of Fig. 9·10. The transfer functions and conversion constants for the various units are given in the diagram. The feedback transfer function is obtained readily through multiplication of the component transfer functions.

$$KG(s) = \frac{\theta_o(s)}{E(s)} = \frac{k_p k_a k_m}{s(Js + f)(Ls + R)}. \tag{66}$$

Hence, the error transfer function is given by

$$\frac{E(s)}{\theta_i(s)} = \frac{1}{1 + KG(s)} = \frac{s(Js + f)(Ls + R)}{s(Js + f)(Ls + R) + k_p k_a k_m}. \tag{67}$$

To proceed with the formal procedure outlined earlier for solution of the error equation, the expressions in numerator and denominator would be multiplied out to give polynomials. The term $\theta_i(s)$ would be transferred to the right side, and some standard input function such as $1/s$ or $1/s^2$

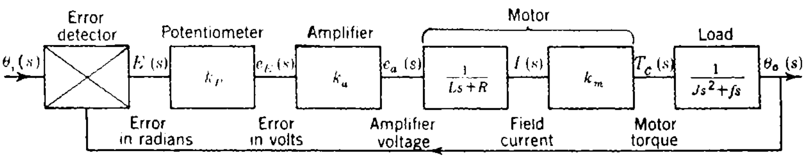


FIG. 9·10.—Block diagram of system containing two energy-storage elements.

substituted for it. The characteristic equation, which would now be of degree three, would be solved for its roots, and so on. The point of special interest here is that the characteristic equation would be of degree three. If more energy storage units were present, the order would be still higher. As is well known, solution of algebraic equations of degree three and higher is somewhat more difficult, the degree of difficulty tending to increase disproportionally with the degree of the equation. It is the necessity of solving for the roots of such higher-degree algebraic equations that makes it laborious to carry to completion the transient solution of the error equation of higher-order functions. The solution is carried through in exactly the same way already outlined for the second-order transfer function. To solve the higher-degree equation for its roots, a number of methods are available such as Graeffe's root-squaring method, Cardan's method, and synthetic division used in conjunction with graphical methods. The reader can find these methods and others described in standard algebraic and engineering texts.[1]

[1] Representative references are: E. Smith, *Automatic Control Engineering*, McGraw-Hill, New York, 1944, pp. 246–252; R. S. Burington, *Handbook of Mathematical Tables and Formulas*, Handbook Publishers, Sandusky, Ohio, 1941; M. Merriman and R. S. Woodward, *Higher Mathematics*, Wiley, New York, 1886; R. E. Doherty and E. G.

Where one finds it necessary only occasionally to solve higher-degree equations, it will probably be found easiest to rely on a comparatively simple method such as synthetic division rather than to go to the labor of learning the more complex methods.

At best, however, such solutions are laborious, and it is found that the transient solutions even when obtained are not so useful for design purposes as might be expected on the basis of treatment of servo systems with second-order transfer functions. Attempts have consequently been made, on the one hand, to develop methods that avoid or reduce the labor involved in solution of higher-degree algebraic equations and that will, on the other hand, fit in more directly with problems of design. Some of the methods concerned primarily with the first problem are considered in the following section. Methods intended to bypass the difficulties associated with the transient solution as well as to facilitate design are represented by the frequency approach to be considered in the next chapter.

9·9. Short-cut Methods and Part Solutions of System Equation.—The previous section has described procedures for determining the complete response of a servo system to applied excitation functions. This response may be formulated in terms of either the output or the error as a function of time, but either form can be readily converted into the other. The nature of the response will be given by an analytic expression that will include both transient and steady-state terms. The transient terms are those which approach zero as t approaches infinity, and the steady-state terms those which remain constant as t approaches infinity. From the analytic representation of the complete response or from the corresponding graphs, one can answer any questions that may be raised concerning special features of the response.

Problems arise, however, in which it is not necessary to have complete details concerning all features of the response. One may, for example, need to know only whether or not a particular type of time function, such as an oscillatory phenomenon, is present or whether the transients are substantially over in a certain time interval. The question arises,

Keller, *Mathematics of Modern Engineering*, Vol. I, Wiley, New York, 1936; L. E. Dickson, *First Course in Theory of Equations*, Wiley, New York, 1922; J. B. Scarborough, *Numerical Mathematical Analyses*, The Johns Hopkins Press, Baltimore, 1930; E. T. Whittaker and G. Robinson, *The Calculus of Observations*, Blackie, Glasgow, 1929, pp. 106–118; and T. von Karman and M. A. Biot, *Mathematical Methods in Engineering*, McGraw-Hill, New York, 1940. Papers bearing particularly on the determination of complex roots are those of S. Liu, "Method of Successive Approximations for Evaluating the Real and Complex Roots of Higher Order Polynomials," *Jour. Math. and Physics*, **20**, No. 3, August 1941; and H. S. Sharp, "Comparison of Methods for Evaluating the Complex Roots of Quartic Equations," *Jour. Math. and Physics*, **20**, No. 3, August 1941.

therefore, whether or not such limited items of information can be obtained by less laborious procedures than those described, i.e., can we obtain special types of information without bothering to obtain the rest. In the present section are collected a number of methods that have such limited objectives. In certain problems they may be capable of giving us all the information that we require; in others they may help by giving us a relatively quick estimate of certain features of the response as a preliminary to the carrying through of the complete solution. The following topics will be considered: (1) determination of steady-state response, (2) inferences from roots of the characteristic equation, (3) inferences from coefficients of the characteristic equation, and (4) graphical methods and higher-degree equations.

Determination of Steady-state Errors.—Since the steady-state errors represent important performance properties, it is often of interest to be able to determine their magnitude without the necessity of obtaining the complete solution. This result may be achieved through application of one of the special theorems of Laplace transform theory, the final value theorem. It states that

$$\lim_{t\to\infty} F(t) = \lim_{s\to 0} sF(s).$$

It is applicable when the function $F(s)$ has no poles on the $j\omega$ axis or in the right half plane.[1]

In applying this theorem to determination of steady-state errors, one starts with the error equation of a given system and substitutes the appropriate function for $\theta_i\,(s)$ such as K/s if the displacement error is to be determined, N/s^2 if the velocity error, and so on. (Here K and N are constants that represent the magnitude of a step function and the slope of a ramp function respectively.) At this point, one merely multiplies $E(s)$ by s and evaluates the limit of $sE(s)$ as $s \to 0$. The procedure is illustrated below for the error equation of a proportional servo system. It is assumed that the *velocity* error is to be determined; hence $\theta_i\,(s)$ is taken as N/s^2.

$$E(s) = \frac{Js^2 + fs}{Js^2 + fs + k_0}\,\theta_i\,(s) = \frac{Js^2 + fs}{Js^2 + fs + k_0}\,\frac{N}{s^2} \tag{68}$$

$$= \frac{(Js + f)N}{s(Js^2 + fs + k_0)}. \tag{69}$$

$$\lim_{t\to\infty} E(t) = \lim_{s\to 0} \frac{(Js + f)N}{(Js^2 + fs + k_0)} = \frac{f}{k_0}\,N \tag{70}$$

$$= \frac{1}{k_v}\,N \qquad \text{if } k_v = \frac{k_0}{f},$$

[1] Gardner and Barnes, *Transients in Linear Systems*, Vol. I, Wiley, New York, 1942, p. 265.

the so-called "velocity error constant." If the *displacement steady-state error* is wanted, K/s is substituted for $\theta_1(s)$ and it is found that

$$E(s) = \frac{(Js + f)K}{(Js^2 + fs + k_0)},$$

and

$$\lim_{t \to \infty} E(t) = \lim_{s \to 0} \frac{s(Js + f)K}{Js^2 + fs + k_0} = \frac{0}{k_0} = 0.$$

Hence, in the case of the proportional servo system with an inertia-viscous friction load, the steady-state displacement error is zero, and the velocity lag error is N/k_v, or the slope of the input velocity function divided by the velocity error constant.

It is of interest to compare the procedure involved in application of the final value theorem and that involved in evaluation of the steady-state term when one obtains the complete solution. Let us review the steps carried out in finding the complete solution of Eq. (69).

$$E(s) = \frac{(Js + f)N}{s(Js^2 + fs + k_0)} = \frac{A}{s} + \frac{B}{s + s_2} + \frac{C}{s + s_3}, \tag{71}$$

where A, B, and C represent constants that are still to be determined and $-s_2$ and $-s_3$ are the roots of the quadratic factor in the characteristic function. It will be recalled that a term such as A/s, corresponding to a first-order pole at the origin, has the constant A as its inverse transform. If the remaining terms do not have poles on the $j\omega$ axis or in the right half plane, then A will be the value of the steady-state error. Poles on the $j\omega$ axis would indicate a sinusoid of constant amplitude; poles in the right half-plane, a time function (either oscillatory or exponential) of progressively increasing magnitude. First-order poles in the left half plane, however, signify decreasing exponentials or damped oscillations, which will tend to zero as $t \to \infty$. Consequently, if the real part of roots $-s_2$ and $-s_3$ are negative, this latter situation holds, and the second and third terms in Eq. (71) can be neglected, since they give rise in the time solution to transient terms. The steady-state error is thus given by A. If A is evaluated by the procedure described in Sec. 9·8, it is found that

$$\begin{aligned} A &= s\left[\frac{(Js + f)N}{s(Js^2 + fs + k_0)}\right]_{s=0} \\ &= \frac{f}{k_0} N. \end{aligned} \tag{72}$$

We thus find, by comparison of Eq. (72) with Eq. (70), that the procedure prescribed for evaluating the constant A in a partial fraction expansion is identical with that called for by the final value theorem. Hence, application of the theorem is equivalent to carrying out the inverse transformation of the error equation and discarding all terms in the partial fraction expansion other than those with a first-order pole at $s = 0$.

Inferences from Roots of Characteristic Equation.—In obtaining the complete solution of the error equation, one essential step, as a preliminary to the partial fraction expansion, was determination of the roots of the characteristic equation. These roots were then represented in the denominators of the various partial fraction terms [see Eq. (31)] and subsequently determined the form of the time functions that appear upon carrying out the inverse transformation. It is thus evident that the nature of the roots will determine the form of the various terms in the time solution. Hence, if one is interested only in the form of the time solution it is not necessary to carry out all the steps that would be required for the complete solution.

The specific correlations that hold between the character of the roots and the time functions are probably most clearly exhibited by diagrams such as those of Table 9·1*b*, which shows the relation of the poles of the function and the corresponding time function. Specification of the location of a pole is equivalent to plotting a root of the characteristic equation in the complex plane. Wherever a complex root occurs, there is also a conjugate root, and hence poles not on the real axis occur in conjugate pairs.

The chief relations that hold between the position of the poles and the form of the time functions may be summarized as follows:

1. Consider first how the time function changes as the poles move from the right to left side of the complex plane. Poles in the right half plane indicate an unstable system, signifying time functions that grow progressively in magnitude; poles on the imaginary axis signify steady-state functions, a constant if the pole is at the origin and a sinusoid of constant amplitude if the poles are on both sides of the real axis; poles in the left half-plane signify transient terms, decreasing exponentially with time. If the poles are paired, lying above and below the real axis, the time function is a sinusoid with an exponential envelope. If the pole lies on the real axis, the time function is simply an exponential (the carrier sinusoid may perhaps be regarded here as of zero frequency). The distance of the pole from the imaginary axis corresponds to the rate of decrease or increase of the time function, *i.e.*, to the slope of the exponential—the greater the distance the greater the slope. In the special case in which the poles lie on the imaginary axis, the real part of the root is zero, and consequently the slope of the exponential curve or envelope is zero.

2. Consider second how the time function changes as a pair of poles, corresponding to conjugate complex roots, moves closer and closer to the real axis. The imaginary part of the root, represented graphically by the distance of the poles from the real axis, is equal to the angular frequency of the sinusoid. Therefore, as this distance decreases, the frequency of the oscillation decreases progressively to zero.

The relations just summarized all relate to first-order poles, corresponding to any given root occurring singly. Similar relations, corresponding to higher-order poles, may be readily derived by inspection of a suitable table of transform pairs. It will be found helpful to visualize or construct a graphical plot of the poles while verbally formulating the relationships.

Inferences from the Coefficients of the Characteristic Equation.—If the characteristic equation is of the third or fourth degree or greater, determination of the roots is laborious as already pointed out. Hence efforts have been made to find clues to the nature of the time response at a still earlier stage in the process of solution. With respect to the stability of the system (unstable systems being represented by time functions of increasing magnitude), Routh found such an index in the relations between the signs and magnitudes of the coefficients of the characteristic equation. One *necessary* condition for stability is that all the terms in the equation be of like sign. Even if this condition is met, however, the system may still be unstable. To determine the supplementary conditions required for stability, a determinant must be set up based on the relative magnitudes and signs of the coefficients. A clear description of the method, with examples, may be found in Gardner and Barnes. Routh's criterion rests ultimately on the same basis as that of using the roots of the characteristic equation, since the character of the roots depends on the relations existing between the coefficients of the original equation. Relations of this sort will be familiar to the reader acquainted with the theory of equations.

Graphical Methods and Higher-degree Equations.—Frequent reference has been made above to the difficulty involved in solving system equations with higher-degree characteristic equations. This circumstance has led to attempts to develop easier methods for determining the nature of the transient response. A way out has been sought through the use of graphical procedures based on the use of relational parameters and nondimensional response curves. Two different though related types of approach may be distinguished: (1) the development of special charts for finding the roots of specific equations of higher degree, and (2) a general procedure of factoring an algebraic equation of any degree greater than the second into a series of factors of first or second degree. Both approaches represent extensions of the nondimensional methods described in Sec. 9·8, originally developed by Draper, Brown, and their students.

The first approach has been followed by Weiss, Liu, and Evans[1] through the provision of special charts for determining the roots of third- and fourth-degree equations. Weiss' charts, for finding the roots of the cubic equation, are based on the writing of the cubic in the form

$$p^3 + 2\zeta_c\omega_n p^2 + \omega_n^2 p + S\omega_n^3 = 0, \tag{73}$$

with p representing the variable and the other terms specially defined relational parameters. Liu and Evans also developed cubic charts, but on the basis of the cubic written in terms of a different set of parameters. They represent the cubic as the product of a linear and quadratic factor, as in Eq. (74), and define their parameters ω_{nq} and ζ_q as the undamped natural period and the damping ratio, respectively, of the quadratic factor.[2] The factor ξ is selected so that $-\xi\omega_{nq}$ represents the single real root of the cubic.

$$(p + \xi\omega_{nq})(p^2 + 2\zeta_q\omega_{nq}p + \omega_{nq}^2) = 0. \tag{74}$$

The merit of this choice of relational parameters lies in the fact that the nondimensional response curves already developed for the quadratic (see Sec. 9·8) can now be used in determining the characteristics of the oscillatory component of the cubic.[3] The parameters ζ_c and ω_n of Weiss' equation do not, on the contrary, represent a damping ratio and undamped natural frequency but are defined simply in terms of the coefficients of the original cubic equation. In Brown's opinion, Liu's charts have

[1] H. K. Weiss, "Constant Speed Control Theory," *Jour. Aero. Sci.*, **6**, No. 4, February 1939; Y. J. Liu, *Servomechanisms: Charts for Verifying Their Stability and for Finding the Roots of Their Third and Fourth Degree Characteristic Equations*, privately printed by Department of Electrical Engineering, Massachusetts Institute of Technology, 1941; L. W. Evans, *Solution of the Cubic Equation and the Cubic Charts*, privately printed by Department of Electrical Engineering, Massachusetts Institute of Technology, 1941. A stability chart for the cubic, based on Liu's chart, may also be found in E. S. Smith, *op. cit.*, p. 242.

[2] For comments on these two procedures for developing cubic charts, see G. S. Brown, *Transient Behavior and Design of Servomechanisms*, privately printed by M.I.T., Department of Electrical Engineering, 1941 and 1943, pp. 30*ff.*; and G. S. Brown, and A. C. Hall, *Dynamic Behavior and Design of Servomechanisms*, preprint ASME meetings, November 1945, p. 18.

[3] The possibility of dealing with the cubic in this way follows from the simple inference that a cubic must contain at least one real root. Since a third-degree equation must have three roots, and since complex roots always occur in conjugate pairs, one of the three roots must of necessity be real. It is the one represented in the linear factor $(p + \xi\omega_{nq})$. The other two roots may be either real or complex depending on the value of ζ_q. The question of which they are is left open by incorporating these two roots as the roots of a quadratic.

proved somewhat more useful in servo-system analysis than have the charts of Weiss.

Liu's charts for obtaining the roots of the fourth-degree equation are based on the procedure of factoring the fourth-degree polynomial into two quadratic factors, each of which may have two roots which can be real or complex. This method of factoring is shown in Eq. (75), taken from Liu.

$$(\lambda^2 + 2\zeta_1\omega_{r_1}\lambda + \omega_{r_1}^2)(\lambda^2 + 2\zeta_2\omega_{r_2}\lambda + \omega_{r_2}^2) = 0, \qquad (75)$$

where λ = the independent variable,
ζ_1 = the damping ratio of component (or factor) 1,
ζ_2 = the damping ratio of component (or factor) 2,
ω_{r_1} = the natural frequency of component 1,
ω_{r_2} = the natural frequency of component 2.

The treatment of the quartic thus follows along the same lines as that developed for the cubic. Since there are four roots, they may be regarded as grouped in pairs as the roots of two quadratics. The parameters ζ, a damping ratio, and ω_r, a natural frequency, have the same significance as indicated in the discussion of the second-degree characteristic equation.

The second approach mentioned above represents an extension of Liu's method of factoring third- and fourth-degree equations. A polynomial of *any order* is represented as the product of a series of quadratics if the polynomial is of even order. There is an additional linear factor if the order is odd. The reason for this is exactly the same as that given for the presence of a linear factor in the case of a cubic equation. In analytical form, the factoring of a higher-degree equation in this way is shown by Eq. (75), based on Brown and Hall.[1]

$$\begin{aligned} p^n + bp^{n-1} + cp^{n-2} + \cdots \\ = (p^2 + 2\zeta_a\omega_{na}p + \omega_{na}^2)(p^2 + 2\zeta_b\omega_{nb}p + \omega_{nb}^2) \cdots (p + a). \qquad (75) \end{aligned}$$

As explained by Brown and Hall:[2]

> Each quadratic factor contributes to a mode of oscillation in the solution having damping ratios ζ_a, ζ_b, ζ_c, and undamped natural frequency ω_{na}, ω_{nb}, ω_{nc}, and so on. Then by the principle of linear superposition the servomechanism response is the sum of the responses attributed to the specific modes a, b, c, etc. Thus for each component of the error response the duration of the transient is given qualitatively by reference to the types of solution given . . . for simple quadratics, and the magnitude can often be approximated from the observation that the higher the magnitude of the root the smaller the coefficient of the time solution involving that root.

[1] For further discussion of this method of treating higher degree characteristic equations, see Brown and Hall, *op. cit.*, p. 22.

[2] *Ibid.*, p. 22.

9·10. Summary.—The present chapter has introduced the general subject of the theoretical study of servomechanisms. The three basic questions in servo theory are: (1) how does the system perform; (2) how does its performance compare with specifications; and (3) if it fails to meet specifications, how can it be modified so that it will do so. Preliminary concepts and methods discussed have included: transformation and operational methods, particularly the Laplace transform methods; transfer functions; generalized block diagrams; and standard forms of input functions. The setting up and the solving of system equations have been treated, together with short cut methods of value both in obtaining complete solutions rapidly and in obtaining partial solutions which answer specific questions (e.g., stability). The emphasis of the present chapter has been on transient solutions.

The next two chapters will consider the steady-state frequency analysis approach, evaluation and correction of system performance, and a number of special problems such as nonlinearity and change of gain.

CHAPTER 10

SERVO THEORY: FREQUENCY ANALYSIS

BY G. L. KREEZER

10·1. Introduction.—A second approach to the analysis and design of servo systems is based on the steady-state response of the system to sinusoidal inputs. This method of analysis has two chief merits compared with the transient method: (1) It is less laborious to apply when the system becomes relatively complicated and the degree of the characteristic equation exceeds two or three, and (2) it lends itself more directly to the development of design procedures for improving the system. The method depends essentially on the construction and interpretation of graphs representing the steady-state response of the system to sinusoidal inputs covering an appropriate frequency range. This method will consequently be referred to as the frequency approach.

The steady-state response to a sinusoidal input of any particular frequency will be a sinusoid of the same frequency but generally differing in amplitude and phase. Hence the response at a given frequency can be completely defined by giving its amplitude and phase relative to that of the input signal. These relations can be concisely represented symbolically.[1]

If $\theta_i(t) = A_i \cos(\omega t + \phi_i) = \mathrm{Re}(A_i e^{j\phi_i} e^{j\omega t})$ and

$$\theta_o(t) = A_o \cos(\omega t + \phi_o) = \mathrm{Re}(A_o e^{j\phi_o} e^{j\omega t}),$$

then the relation of $\theta_o(t)$ to $\theta_i(t)$ at any particular frequency ω can be completely specified by means of the ratio $P(j\omega)$ of the rotating vectors used to represent them. The symbol "Re" designates the real part of the complex number or function.

$$P(j\omega) = \frac{A_o e^{j\phi_o} e^{j\omega t}}{A_i e^{j\phi_i} e^{j\omega t}} = \frac{A_o}{A_i} e^{j(\phi_o - \phi_i)}. \tag{1}$$

The value of this ratio, which may be designated as the transfer ratio, is thus a complex number with A_o/A_i constituting the modulus or absolute value and $(\phi_o - \phi_i)$ the phase angle. Like any complex number, it can be regarded as represented by a vector drawn from the origin of the

[1] The method of representation follows closely that of Hall. The terminology here differs from Hall, however, in that the term "servo system transfer function" is not limited to the feedback transfer function.

complex plane. Alternatively, it can be specified by giving its real and imaginary parts. The modulus A_o/A_i gives the ratio of output to input amplitude, and $(\phi_o - \phi_i)$ gives the magnitude of phase difference between output and input. A negative sign will indicate that θ_o lags behind θ_i, and a positive sign that it leads. Now if ω is allowed to take on a range of different values, we obtain a specific amplitude ratio and a specific phase difference for each value of ω. The set of values of the amplitude ratio corresponding to different values of ω may be designated as the *amplitude*[1] or *gain function*, and the set of values for the phase as the *phase function*. The amplitude and phase functions together constitute the *frequency response* of the system. These two frequency-response functions can be plotted as graphs in a number of different ways which will be described below.

Any such graphical representation may be regarded as a means for representing the characteristics of the system, and the ensuing interpretation of the graphs a procedure for determining performance or response characteristics. The steps involved in thus determining system performance by the frequency approach consist in (1) determination of frequency-response data for the system, (2) plotting the data by one or more of the available methods, and (3) interpretation of the graphs to give performance characteristics of the system. Each of these steps will be considered in turn in the following sections. A final section will summarize the kind of operations that may be performed on the various types of frequency diagram.

In the section on preliminary concepts, three different kinds of transfer function have been defined: the feedback transfer function $\theta_o(s)/E(s)$, the output transfer function $\theta_o(s)/\theta_i(s)$, and the error transfer function $E(s)/\theta_i(s)$. As already pointed out, the last two functions are uniquely determined if the feedback transfer function is known. Although the different ways of plotting frequency-response curves might formally be used for any one of the three types of transfer function, the frequency approach relates principally to the feedback transfer function. Unless otherwise specified, the sections to follow will relate to it. Some use is also made of the output-transfer function as a basis for the development of certain principles of interpretation. The error transfer function, which plays so prominent a role in the transient analysis, is not used at all in the frequency method.

10·2. Determination of Frequency-response Data.—Since the application of the frequency approach is based on the use of graphs of the frequency response of the system, the first need is to obtain the data

[1] This use of the term amplitude, in accordance with conventional a-c terminology, should not be confused with the occasional use of amplitude in mathematics to designate the argument or phase angle of a complex number.

corresponding to the amplitude and phase functions. A number of alternative methods are available. In case the feedback transfer function of the system is known analytically, the amplitude and phase functions can be computed from the transfer function. A second computational procedure has been described by Bode (see below), which permits either the phase or amplitude function to be computed from the other function. Finally, if the physical system is available, the necessary data may be obtained empirically.

Computation from the Transfer Function.—From the definition given of the transfer function, it follows that for a known transfer function and any given input function, the value of the output response may be computed. Thus we may assume the input to be a sinusoidal wave and represent it by the appropriate operational expression (as given, for example, by transform pairs *b* or *c* of Table 9·11), compute the steady-state part of the output response, and thus determine the amplitude and phase relations of output to input. It turns out that the same result can be obtained more easily by starting with the transfer function of the system $KG(s)$, substituting $j\omega$ for s to give the frequency transfer function $KG(j\omega)$, and finally substituting an appropriate range of specific values for the angular frequency ω, to obtain the amplitude and phase functions. At any particular value of ω, the transfer function will be a complex number that can be specified in terms of either its modulus and phase angle or its real and imaginary parts. The set of values of amplitude and phase thus determined provides the necessary data for plotting amplitude and phase functions.

The relation of this procedure to the concepts of complex function theory is of interest. The function $KG(s)$ is a function of the complex variable s. The nature of the function can be examined by means of corresponding plots on two complex planes, on the complex s-plane to show a given set of values of s and on the complex $KG(s)$-plane to show the corresponding values of $KG(s)$. If it is assumed that s takes on the series of values extending from $-j\infty$ to $+j\infty$, this set of points is represented in the s plane by the $j\omega$, or imaginary axis. The $KG(j\omega)$ function defines the corresponding set of values in the $KG(s)$ plane. When plotted, the resultant curve constitutes the conformal map of the $j\omega$ line on the $KG(s)$ plane. As will be apparent later, this conformal plot is identical with the transfer locus, which plays an important part in the analysis of feedback systems by means of the frequency approach.

Phase Function and Amplitude Function Computed from Each Other.—Bode, in his discussion of design procedures for feedback amplifiers, has emphasized the fact that in a minimal phase system the amplitude and phase functions are not independent. When one is specified, the other is thereby determined. Mathematical investigations bearing on this relationship had been made by a number of previous workers, though in different contexts.[1] Bode has also worked out procedures to facilitate computation of the one function when the other is given or when only parts of each function are given for different parts of the frequency range. These methods are of special value in cases in which a physical system has been set up but the feedback transfer function is not known analytically. In such cases the amplitude function may usually be determined quite readily by direct measurement, but empirical measurement of the phase response is often difficult. Bode's methods make it possible to dispense with physical measurement of the phase response and to compute it from the amplitude function. For a description of these methods, Bode should be consulted.[2] The reciprocal dependence of the two functions has also been utilized by Bode to develop criteria of basic importance in the design of feedback amplifiers.[3] These same criteria are also relevant in the design of servo systems, since servo systems are exact analogues of feedback amplifiers.

Empirical Determination of Frequency-response Data.—It is also possible to obtain the data corresponding to amplitude- and phase-frequency functions by means of direct physical measurement. If a sinusoidal oscillation of constant amplitude is introduced at the input of the servo system, then within a relatively short interval, steady-state conditions may be assumed to be operative and all points of the system will show sinusoidal oscillations of the same frequency but possibly differing in amplitude and phase from each other. If the amplitude and phase of this oscillation for the error $E(t)$ and the output $\theta_o(t)$ are measured by appropriate experimental methods,[4] the amplitude ratio and phase difference can be computed, the amplitude and phase of the error being taken as reference. If the input signal is made to vary in frequency from zero through an appropriate frequency range, the amplitude and phase relations can be determined as a function of frequency.

If it is experimentally possible to measure the output response and error response directly, then it does not matter in theory whether the feedback loop is open or closed. For although the properties of the oscillation at a given point will differ for these two conditions, the procedure of measurement assures our obtaining the ratio of the vectors representing $\theta_o(t)$ and $E(t)$. This ratio must always be the same for a

[1] See, e.g., Y. W. Lee, "Synthesis of Electric Networks by Fourier Transformation of Laguerre's Functions," *Jour. Math. and Physics*, **11**, 83–113, June 1932.

[2] H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945, pp. 303–359; see also Sec. 10·3 of this volume.

[3] H. W. Bode, *op. cit.*, pp. 451–488; "Relations between Attenuation and Phase in Feedback Amplifier Design," *Bell System Tech. Jour.*, **19**, 421–454, July 1940; U.S. Patent 2123178, July 12, 1938.

[4] See Chap. 13.

given frequency, since it depends only on the parameters of the part of the system lying between the "error point" and the "output point" (see Fig. 9·3). One must be sure, however, that the error, defined as $\theta_i(t) - \theta_o(t)$, is actually being measured, for in many systems this quantity may not be accessible to physical measurement but must be inferred from measurements made at some adjacent point of the system. In such instances, if the amplitude and phase of the error function cannot be computed by a mathematical relation, it may be necessary to make measurements with the feedback loop open at a convenient point. Then the vector ratio of the a-c signal at the end of the loop to the signal fed in at the beginning of the loop will correspond to the desired ratio $\theta_o(j\omega)/E(j\omega)$ at the given value of ω (except for a possible reversal of sign).[1] If the signal is actually fed in at the θ_i point and measured at the θ_o point, then the θ_o/θ_i ratio will be identical with the θ_o/E ratio, since with loop open or broken, the error and input sinusoids are the same. For in this instance,

$$E(t) = \theta_i(t) - \theta_o(t) = \theta_i(t) - 0. \tag{2}$$

In some systems, however, opening the loop may make measurements difficult due to system instability occasioned by removal of the corrective influence of the negative feedback link. Experimental techniques and precautions necessary for actually measuring amplitude and phase relationships are considered in Chap. 13.

10·3. Graphical Plots of the Frequency Response.—The utilization of the frequency method of analysis hinges on the fact that important performance properties of the system can be determined from graphs of the frequency functions. In the present section different ways of plotting these functions are described. The interpretation of the graphs in terms of performance properties will follow in the next section. As a preliminary to the methods described below, it is assumed that the value of the complex number representing the transfer ratio of output to input has been obtained for an appropriate range of values of ω.

Transfer-locus Plot in the Complex Plane.—For any given value of ω, a point may be plotted in the complex plane for each pair of values specifying the value of the complex number representing $\theta_o(j\omega)/E(j\omega)$. The real and imaginary parts may be used to find the point or the modulus and phase angle, as polar coordinates. The line connecting all of these points, for various values of ω, is designated as the *transfer locus* (following Hall's terminology) or the *Nyquist diagram*. At representative points along the locus, the value of ω may be specified to indicate the corre-

[1] In making such measurements, it is, of course, necessary that impedance relations not be allowed to change when the loop is broken and an a-c source and measuring instruments inserted at the place where the loop is broken. See Chap. 13.

spondence of different regions of the locus with the frequency scale. The vector drawn from the origin to any point on the locus represents the transfer ratio of output to input at that frequency. An example of this type of graph is shown in Fig. 10·1. Reference to Sec. 10·2 will show that the transfer locus is identical with the conformal mapping on the complex $\theta_o(s)/E(s)$ plane of the line $s = j\omega$.

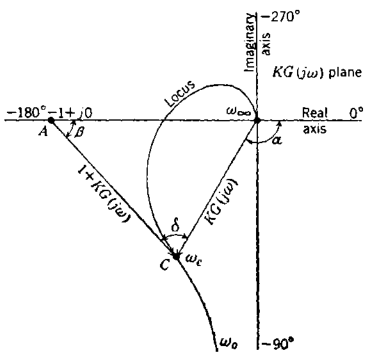


FIG. 10·1.—Graphical determination of system and error transfer functions.

Decibel vs. Log Frequency Diagrams.[1]—If the frequency-response data corresponding to a set of values of ω is given in terms of the modulus $|\theta_o(j\omega)|/|E(j\omega)|$ and the phase lag ϕ, then the relation of these quantities to frequency may be shown by plotting two separate curves in rectangular coordinates, the modulus, or gain, in decibels against logarithm of the frequency and the phase against logarithm of the frequency. In some cases it is satisfactory to plot the two curves on separate grids; in other

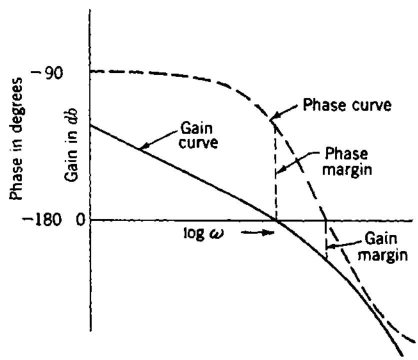


FIG. 10·2.—Nyquist's stability criterion as applied in decibels vs. log frequency method of plotting feedback transfer function.

cases, when the relationship between the two is of paramount interest, as in the application of the Nyquist criterion, then it is more useful to

[1] A reference to the literature dealing with frequency-response curves or decibel vs. log frequency diagrams may be confusing at times, since a considerable variety of terms may be used to refer to the curve showing the dependence of the *absolute value* of transfer ratio on frequency. This is due to the variety of terms that can be applied to this magnitude. Among the terms that have been used are amplitude response, gain, attenuation, and decibel. Although these terms all refer to the same transfer ratio, they may represent it in different units or with different signs.

plot them on the same grid. In such cases, for reasons that will be apparent in Sec. 10·4, it is helpful to let the x-axis represent both the 0-db level and a phase lag of 180°. An example of such a plot is given in Fig. 10·2. In this figure, the two separate scales for gain and phase are indicated along the y-axis.

It is of interest to note that this type of graph really represents a plot of the logarithm of the transfer ratio against the logarithm of ω, in the form of two separate curves. For if

$$\frac{\theta_o(j\omega)}{E(j\omega)} = \left|\frac{\theta_o(j\omega)}{E(j\omega)}\right| e^{j\phi(\omega)} = R(\omega)e^{j\phi(\omega)}, \tag{3}$$

where $R(\omega)$ represents the modulus, expressed as a function of ω, and $\phi(\omega)$ specifies the phase angle as a function of ω, then, taking the natural logarithm of each side,

$$\ln \frac{\theta_o(j\omega)}{E(j\omega)} = \ln R(\omega) + j\phi(\omega). \tag{4}$$

We thus have a new complex function in which the ln $R(\omega)$ is the real part and $j\phi(\omega)$ is the imaginary part. Each part may be plotted separately against the logarithm of ω. It has been customary to plot R in decibels (which amounts to a plot of 20 $\log_{10} R$), and ω as $\log_{10} \omega$. Plotting R and ω in these units rather than in terms of the natural logarithms given by Eq. (4) amounts merely to a change in the size of the scale units used in the graph and does not alter the logarithmic relationships between the variables.

Decibel vs. Log Frequency Diagrams: Approximate Curves.—In place of the exact decibel log ω curves, approximate curves can be used that represent the data with sufficient accuracy and have certain special advantages.[1] These advantages include elimination of laborious computation, the provision of valuable indices of system accuracy, and the direct graphical representation of the time constants of the dynamic units of the system.

For purposes of illustration, let us again use the transfer function of the proportional servo system, with a load member involving inertia and viscous friction, as given by Eq. (5). It is convenient here to sub-

[1] Descriptions of or references to this approximation method of plotting decibel log ω diagrams have been given by E. B. Ferrell, "The Servo Problem as a Transmission Problem," *Proc. IRE*, **33**, 763–767, November 1945; L. A. McColl, *Fundamental Theory of Servomechanisms*, Van Nostrand, New York, 1945, pp. 45–48; N. B. Nichols, see following reference; D. P. Campbell, "A Discussion of the Db-log Frequency Methods of Analysis and Synthesis of Automatic Control System Behavior," based on lecture by N. B. Nichols, Massachusetts Institute of Technology, Dec. 21, 1945; H. Lauer, R. Lesnick, and L. E. Matson, *Servomechanism Fundamentals*, Chap. 9, McGraw-Hill, New York, 1947. See also Vol. 25, Radiation Laboratory Series.

stitute predimensional parameters k_v and T for the physical parameters to give the transfer function in the form of Eq. (6) [see Eq. (44), sec. 9·8].

$$KG(s) = \frac{k_0}{Js^2 + fs}. \tag{5}$$

$$KG(s) = \frac{k_v}{Ts^2 + s}. \tag{6}$$

Substituting $j\omega$ for s to give the frequency transfer function, we obtain

$$KG(j\omega) = \frac{-k_v}{\omega(T\omega - j)} = k_v \frac{1}{\omega(T^2\omega^2 + 1)} (-T\omega - j). \tag{7}$$

In determining the *form* of the gain curve of this function, we need consider only the frequency-dependent portion

$$G(j\omega) = \frac{1}{\omega(T^2\omega^2 + 1)} (-T\omega - j). \tag{8}$$

The constant gain factor k_v can have the effect only of changing the position of the whole curve relative to the 0-db axis but cannot change its shape. Adjustment of the position of the curve corresponding to the magnitude of k_v can therefore be made later, after the plot of $G(j\omega)$ has been obtained.

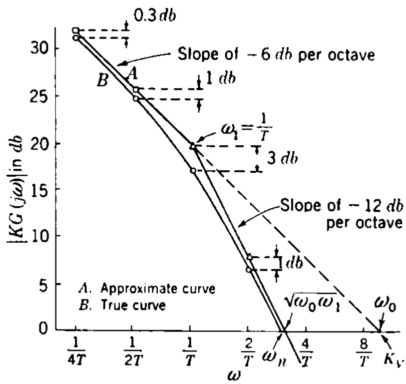


FIG. 10·3.—Gain in decibels vs. log frequency plot of $KG(j\omega)$. This may be taken as a plot of $G(j\omega)$ if the gain constant k or k_v is unity (0 db), in which case $\omega_0 = 1$ radian per second.

An approximate plot of $G(j\omega)$ in terms of the decibel log ω scheme can now be made very simply.[1] The steps involved in plotting the gain curve, $|G(j\omega)|_{db}$ are:[2] (1) Plot the point corresponding to a gain of 3 db plus $|G(j\omega)|_{db}$ at $\omega = 1/T$;[3] (2) to the left of this point, draw a straight line with a slope of -6 db per octave; (3) to the right of the point, draw another straight line with a slope of -12 db per octave.

[1] In making such plots either log-log paper or semilog paper may be used depending on whether gain is plotted in units of decibels or in terms of absolute ratio. It is often helpful to use log-log paper and show the gain scale in terms of both the absolute ratio $|G(j\omega)|$ and decibels. Then one may read off the gain or plot it in either type of unit.

[2] The procedure for plotting the phase curve is given at the end of this section, following complete discussion of the gain curve.

[3] Note somewhat easier procedure for plotting low-frequency asymptote proposed later in this section.

This curve constitutes our approximation to the gain function of Eq. (8). It will be referred to as an asymptotic decibel log ω plot, since the straight lines are asymptotes to the exact curve. Figure 10·3 can be regarded as showing the results of this procedure if the constants K and k_v shown in the figure are understood to equal 1.

What is the justification for the above procedure? In approximating the true curve in this way, four assumptions are involved: (1) that the straight line of -6 db per octave slope can be used to represent the gain curve for values of $\omega \ll 1/T$, (2) that the line of -12 db slope can be used to represent the curve for $\omega \gg 1/T$, (3) that their intersection will occur at $\omega = 1/T$, and (4) that the errors of approximation are sufficiently small at all points to be considered negligible. The maximum value of this error will be 3 db, and it will occur at $\omega = 1/T$.

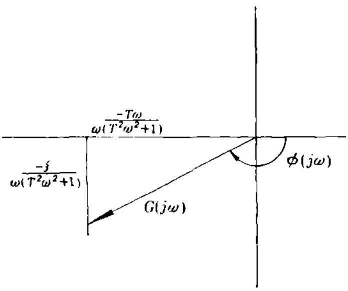


FIG. 10·4.—Computation of modulus and phase angle of $G(j\omega)$.

Justification for these assumptions can be obtained quite readily from a consideration of Eq. (8). From this equation, let us find the value of the modulus $|G(j\omega)|$ and the argument or phase function $\phi(j\omega)$. By means of the conventional procedure for finding the value of the modulus and phase angle of a complex number, we find (see Fig. 10·4)

$$|G(j\omega)| = \frac{1}{\omega\sqrt{T^2\omega^2 + 1}} \tag{9}$$

$$\begin{aligned} \phi(j\omega) &= -180° + \tan^{-1}\frac{1}{T\omega} \\ &= -180° + (90° - \tan^{-1} T\omega) \\ &= -90° - \tan^{-1} T\omega. \end{aligned} \tag{10}$$

Writing Eq. (9) in terms of decibels, we obtain

$$\begin{aligned} |G(j\omega)|_{\text{db}} &= 20 \log \frac{1}{\omega\sqrt{T^2\omega^2 + 1}} \\ &= -20 \log \omega - 20 \log \sqrt{T^2\omega^2 + 1}. \end{aligned} \tag{11}$$

We may now obtain approximate expressions for the gain $|G(j\omega)|_{\text{db}}$ in the two regions in which we are interested.

1. The decibel gain curve as ω approaches zero: When $\omega \ll 1/T$, $T\omega \ll 1$, and $T^2\omega^2 \ll 1$. Therefore

$$\sqrt{T^2\omega^2 + 1} \approx 1,$$

and Eq. (11) becomes

$$|G(j\omega)|_{\text{db}} \approx -20 \log \omega. \tag{12}$$

This equation represents $|G(j\omega)|_{\text{db}}$ as a linear function of $\log \omega$; hence a plot of it against $\log \omega$ will be a straight line. What will be the slope of this straight line in relation to ω? Let us designate as ω_a some value of ω in the region of the curve for which Eq. (12) is valid and compute from Eq. (12) the value of $|G(j\omega)|_{\text{db}}$ at $\omega = \omega_a$, $2\omega_a$, $4\omega_a$, and so on. The results are shown in Table 10·1.

TABLE 10·1

ω	$\|G(j\omega)\|_{\text{db}}$
ω_a	$-20 \log \omega_a$
$2\omega_a$	$-20 \log \omega_a - 6$ db
$4\omega_a$	$-20 \log \omega_a - 12$ db

For each octave increase in ω, the gain line falls by 6 db. This result confirms the assumption made concerning the low-frequency end of the curve.

2. The decibel gain curve for $\omega \gg 1/T$: When $\omega \gg 1/T$, $T\omega \gg 1$, and $T^2\omega^2 \gg 1$. Therefore,

$$\sqrt{T^2\omega^2 + 1} \approx \sqrt{T^2\omega^2} = T\omega.$$

Hence, if $T\omega$ is substituted for $\sqrt{T^2\omega^2 + 1}$ in Eq. (9),

$$|G(j\omega)| = \frac{1}{\omega(T\omega)},$$

and

$$|G(j\omega)|_{\text{db}} \approx -20 \log \omega - 20 \log T\omega. \tag{13}$$

$|G(j\omega)|_{\text{db}}$ is here, too, a linear function of $\log \omega$, since each of the two terms on the right-hand side are linear functions of $\log \omega$. Therefore the decibel curve for $\omega \gg 1/T$ will also be a straight line. To find its slope relative to ω, let ω_b represent a value of $\omega \gg 1/T$, and compute the decibel gain for ω_b, $2\omega_b$, and so on, by substitution for ω in Eq. (13). The results are given in Table 10·2.

TABLE 10·2

ω	$\|G(j\omega)\|_{\text{db}}$
ω_b	$(-20 \log \omega_b - 20 \log T\omega_b)$
$2\omega_b$	$(-20 \log \omega_b - 20 \log T\omega_b) - 40 \log 2$
	$= (-20 \log \omega_b - 20 \log T\omega_b) - 12$ db
$4\omega_b$	$(-20 \log \omega_b - 20 \log T\omega_b) - 24$ db

Thus for every octave increase in ω, the gain curve falls by 12 db.

A somewhat simpler procedure than that given above may be proposed for plotting $|G(j\omega)|_{\text{db}}$ in the interval $0 < \omega < 1/T$, on the basis of the value of ω at which the gain line intersects the 0-db axis. The equation of this line, as given by Eq. (12) is

$$|G(j\omega)|_{\text{db}} = -20 \log \omega. \tag{12}$$

At the point of intersection with the 0-db axis,

$$\begin{aligned} |G(j\omega)|_{\text{db}} &= 0, \\ 0 &= -20 \log \omega, \\ \log \omega &= 0, \end{aligned}$$

Therefore

$$\omega = 1.$$

Hence, this line may be plotted by drawing a straight line of -6 db per octave slope through the point on the 0-db axis corresponding to $\omega = 1$. Then, the point on this line whose abscissa is $\omega = 1/T$ is used as the initial point of the line drawn in the next interval, with a slope of -12 db per octave, given by Eq. (13). These two lines constitute our plot of the asymptotic gain characteristic for Eq. (8).

3. Where will the two straight lines intersect? The fact that the two lines will intersect at $\omega = 1/T$ is shown most easily by substituting $\omega = 1/T$ in the equation for each line, Eqs. (12) and (13). In both cases,

$$|G(j\omega)|_{\text{db}} = -20 \log \frac{1}{T}.$$

Hence the lines must intersect at $\omega = 1/T$.

4. What will be the errors of approximation involved in the use of the asymptotic curves in place of the true curve? Equations for the errors in the low and high regions of the ω scale are readily computed. For low ω, let us subtract the value of gain given by the exact equation (11) from that given by the approximate equation (12). A positive sign for the error will indicate that the approximate curve lies above the exact curve at that point. The symbol ϵ stands for the error of the approximation.

$$\epsilon_{\text{low}\,\omega} = -20 \log \sqrt{T^2\omega^2 + 1}. \tag{14}$$

Similarly for $\omega > 1/T$, we use Eqs. (11) and (13). Then

$$\epsilon_{\text{high}\,\omega} = -20 \log T\omega + 20 \log \sqrt{T^2\omega^2 + 1}. \tag{15}$$

Table 10·3 shows the value of the error computed for various values of ω. The table shows that the error is greatest at the point of intersection of the asymptotes, at $\omega = 1/T$, but even here it is relatively small, equal to only 3 db. As ω varies by octave steps from the point of intersection, the errors become quite negligible. Figure 10·3 shows the relation of the exact and approximate curves, drawn on the basis of these data.

The simple procedure just described for making an approximate plot of the feedback transfer function given by Eq. (6) may now be extended to *transfer functions of any order* and for cases where factors are present in numerator as well as denominator. The steps involved are as follows:

1. The feedback transfer function, if given in terms of physical parameters, is rewritten in terms of the relational parameters k_v and time constants $T_1, T_2, \ldots, T_k, \ldots, T_n$. The various factors are

TABLE 10·3.—ERROR (ϵ) IN $|G(j\omega)|_{\text{db}}$

ω	For $\omega \leqq \dfrac{1}{T}$	For $\omega \geqq \dfrac{1}{T}$
$\dfrac{1}{4T}$	0.3 db	......
$\dfrac{1}{2T}$	1.0 db	......
$\dfrac{1}{T}$	3.0 db	3.0 db
$\dfrac{2}{T}$	......	1.0 db
$\dfrac{4}{T}$	......	0.3 db

arranged in sequence to correspond to progressively decreasing time constants, as $T_1, T_2, \ldots, T_k, \ldots, T_n$. The form of the resultant transfer function, for the general case, is given by Eq. (16).[1] The symbol T_k is used to represent any time constant, whether it appears in numerator or denominator, and T_n the smallest time constant.

$$\frac{\theta_0(s)}{E(s)} = KG(s)$$
$$= \frac{k_v(T_2s + 1)(T_4s + 1)\cdots}{s(T_1s + 1)(T_3s + 1)\cdots(T_ks + 1)\cdots(T_ns + 1)}. \qquad (16)$$

The equation as written does not imply a like number of factors in numerator and denominator. Factors containing s may, in fact, be entirely absent from the numerator, as in cases where the system does not contain phase advance components.

2. Through substitution of $j\omega$ for s, the frequency feedback transfer function is obtained, as in Eq. (17).

[1] The factors in this equation are here represented in the form $(T_ks + 1)$ instead of $\left(s + \dfrac{1}{T_k}\right)$ since the form given first leads readily to the approximations for various factors derived later in this section.

$$G(j\omega) = \frac{(j\omega T_2 + 1)(j\omega T_4 + 1)\cdots}{j\omega(j\omega T_1 + 1)(j\omega T_3 + 1)\cdots(j\omega T_k + 1)\cdots(j\omega T_n + 1)}. \quad (17)$$

In this equation, only the frequency dependent part of the function, $G(j\omega)$, is given, for reasons indicated earlier. Hence k_v is not shown on the right side of the equation.

3. The approximate expression for the gain transfer function $|G(j\omega)|$ can now be written as shown in Column 2 of Table 10·4. The symbol $|\tilde{G}(j\omega)|$ is used to represent the approximate gain function. Table 10·4 shows that $|\tilde{G}(j\omega)|$ differs for different intervals on the ω axis. The location of these intervals is determined by the location, on the ω axis, of the reciprocals of the various time con-

TABLE 10·4.—APPROXIMATE GAIN FUNCTIONS FOR HIGHER-ORDER TRANSFER FUNCTIONS

$$G(j\omega) = \frac{(j\omega T_2 + 1)(j\omega T_4 + 1)\cdots}{j\omega(j\omega T_1 + 1)(j\omega T_3 + 1)\cdots(j\omega T_k + 1)\cdots(j\omega T_n + 1)}$$

Frequency interval	$\lvert\tilde{G}(j\omega)\rvert$ Approx. gain function	$\lvert\tilde{G}(j\omega)\rvert_{db}$	Slope of $\lvert\tilde{G}(j\omega)\rvert_{db}$
$0 < \omega < \frac{1}{T_1}$	$\frac{1}{\omega}$	$-20\log\omega$	-6 db/octave
$\frac{1}{T_1} < \omega < \frac{1}{T_2}$	$\frac{1}{\omega}\frac{1}{\omega T_1}$	$-20\log\omega - 20\log\omega T_1$	-12 db/octave
$\frac{1}{T_2} < \omega < \frac{1}{T_3}$	$\frac{1}{\omega}\frac{1}{\omega T_1}(\omega T_2)$	$-20\log\omega - 20\log\omega T_1 + 20\log\omega T_2$	-6 db/octave
$\frac{1}{T_3} < \omega < \frac{1}{T_4}$	$\frac{1}{\omega}\frac{1}{\omega T_1}(\omega T_2)\frac{1}{\omega T_3}$	$-20\log\omega - 20\log\omega T_1 + 20\log\omega T_2 - 20\log\omega T_3$	-12 db/octave
. . . . . .	. . . . . .	. . . . . .	
$\frac{1}{T_k} < \omega < \frac{1}{T_{k+1}}$	$\frac{1}{\omega}\frac{1}{\omega T_1}(\omega T_2)\frac{1}{\omega T_3}\cdots\frac{1}{\omega T_k}$	$-20\log\omega - 20\log\omega T_1 + 20\log\omega T_2 - 20\log\omega T_3 \cdots - 20\log\omega T_k$	$+6a$ db/octave* $-6b$ db/octave
. . . . . .			

* a equals number of factors in numerator, and b equals number of factors in denominator of $|\tilde{G}(j\omega)|$.

stants. In each succeeding interval an additional factor is present in the function. The process of approximation consists in substituting ω for the $j\omega$ term, ωT_k for each factor of the form $(j\omega T_k + 1)$ where $\omega T_k > 1$, and substituting 1 for each factor where $\omega T_k < 1$. T_k here signifies any time constant. The same type of substitution is made for factors in numerator or denominator. The justification for this procedure will be given below.

4. The approximate gain function is written in decibel form as shown in Column 3 of Table 10·4. The procedure consists simply in taking $20\log_{10}|\tilde{G}(j\omega)|$. Thus, for any given interval

$$\frac{1}{T_k} < \omega < \frac{1}{T_{k+1}}.$$

$$|\tilde{G}(j\omega)|_{db} = 20\log_{10}|\tilde{G}(j\omega)| = -20\log\omega - 20\log\omega T_1 + 20\log\omega T_2 - 20\log\omega T_3 + \cdots - 20\log\omega T_k. \quad (18)$$

The sign is plus for terms in the numerator of $|\tilde{G}(j\omega)|$ and minus for terms in the denominator of $|\tilde{G}(j\omega)|$.

5. $|\tilde{G}(j\omega)|_{db}$ can now be plotted by an extension of the procedure already described for the proportional system. The straight line for the interval $0 < \omega < 1/T_1$ is plotted in exactly the same way as before, by drawing a straight line of slope -6 db/octave through the point (1,0) and locating on the line the point corresponding to $\omega = 1/T_1$. The next linear section of the curve starts from this point and runs to $\omega = 1/T_2$, and so on for each succeeding interval. The slope of the curve in each interval depends on the number of terms present in the approximate gain function $|\tilde{G}(j\omega)|$ in that interval. The line has a slope of 6 db octave for each term (of the form ω or (ωT) present in the approximate gain function. The sign for the slope contributed by each term corresponds to the sign of the logarithmic term. Thus for Eq. (19), the slope of the curve in the interval $1/T_3 < \omega < \infty$ will be $(-6-6+6-6)$ db/octave, or -12 db/octave, as indicated by Eq. (20).

$$G(j\omega) = \frac{(j\omega T_2 + 1)}{j\omega(j\omega T_1 + 1)(j\omega T_3 + 1)}, \tag{19}$$

$$|\tilde{G}(j\omega)|_{db} = -20 \log \omega - 20 \log \omega T_1 + 20 \log \omega T_2 - 20 \log \omega T_3;$$
$$\text{for } \frac{1}{T_3} < \omega < \infty. \tag{20}$$

6. Finally, the whole gain curve is displaced upward by an interval equal to k_v, in decibels.[1]

Some of the steps given above require further elaboration or proof. Let us consider them in the order listed.

The initial step of writing the feedback transfer function in the form shown in Eq. (16) (with the time-constant set of parameters) has already been considered for a system with one time-constant factor (the proportional servo system of Sec. 9·8 [Eq. (9·44)]. Let us see how the procedure is carried out in the more general case. For illustrative purposes, it is sufficient to consider a system with one additional time delay factor such as that represented by Eq. (21) below, previously designated as Eq. 9·66 (Sec. 9·8). The additional delay factor here is provided by the inductance-resistance field control circuit of a d-c motor. The transfer function for the system in terms of physical parameters is

$$KG(s) = \frac{k_p k_a k_m}{s(Js + f)(Ls + R)}, \tag{21}$$

[1] A review of the various steps outlined will indicate that once the procedure has been understood, it can be carried out merely from inspection of Eq. (16) or (17), since these equations contain all the information needed for making the plot, taken in conjunction with the rules just enumerated.

where k_p, k_a, and k_m represent sensitivity or conversion factors of potentiometer, amplifier, and motor respectively. To throw the equation into the form of Eq. (16), the isolated constant in each parenthesis is factored out, and all constants collected in the numerator. Thus

$$KG(s) = \frac{k_p k_a k_m}{sf\left(\frac{J}{f}s + 1\right)R\left(\frac{L}{R}s + 1\right)} = \frac{\frac{k_p k_a k_m}{fR}}{s\left(\frac{J}{f}s + 1\right)\left(\frac{L}{R}s + 1\right)}. \tag{22}$$

Now if the term k_v is substituted for the combination of constant terms in the numerator, T_L for J/f, and T_m for L/R, the equation will be in the required form, as shown in Eq. (23).

$$KG(s) = \frac{k_v}{s(T_L s + 1)(T_m s + 1)}. \tag{23}$$

As defined here, k_v merely signifies the aggregation of constants representing the gain when the transfer function has been manipulated to put it in the form of Eq. (22). What this aggregation includes will depend, of course, on the number and nature of the energy storage components. But it can be shown that k_v, if obtained in this way, will have the dimensions of 1/sec and that it can be regarded as the velocity error constant of the total system (see Sec. 10·4). It cannot be defined in terms of a specific group of physical parameters that is the same for all systems but depends on a group of physical parameters distributed throughout the entire system, in the manner illustrated by the present example.

We may now consider the basis for the procedure proposed in Step 3 for *approximating terms in the gain function* $|G(j\omega)|$. Equation (17) can be rewritten as a product of a series of separate factors.

$$G(j\omega) = \frac{1}{j\omega}\frac{1}{j\omega T_1 + 1}(j\omega T_2 + 1)\frac{1}{j\omega T_3 + 1}\cdots\frac{1}{j\omega T_k + 1}\cdots\frac{1}{j\omega T_n + 1}. \tag{24}$$

Therefore

$$|G(j\omega)| = \left|\frac{1}{j\omega}\right|\left|\frac{1}{j\omega T_1 + 1}\right||j\omega T_2 + 1|\cdots\left|\frac{1}{j\omega T_k + 1}\right|\cdots\left|\frac{1}{j\omega T_n + 1}\right|. \tag{25}$$

In passing from Eq. (24) to Eq. (25) we have merely determined the absolute value of a complex function $G(j\omega)$ by multiplying the absolute values of a series of component complex functions. Thus, at any value of ω, $G(j\omega)$ as well as all the separate factors on the right-hand side are complex numbers. Hence, the absolute value of G will equal the product of the absolute values of all the components. It may be helpful here to think in terms of the familiar conception of a complex number as a vector, with the absolute value being equal to its length. Then

$$|G(j\omega)| = \frac{1}{\omega}\,\frac{1}{\sqrt{(\omega T_1)^2+1}}\,\sqrt{(\omega T_2)^2+1}\,\cdots\,\frac{1}{\sqrt{(\omega T_k)^2+1}}\,\cdots\,\frac{1}{\sqrt{(\omega T_n)^2+1}}. \tag{26}$$

Now, if for any term of the form $\sqrt{(\omega T_k)^2+1}$, $\omega \ll 1/T_k$, then $\omega T_k \ll 1$, $(\omega T_k)^2 \ll 1$, and $\sqrt{(\omega T_k)^2+1} \approx 1$. Thus 1 can be used to approximate the term $\sqrt{(\omega T_k)^2+1}$. This approximation is even closer for all terms of this form still later in the series of terms in Eq. (26), since the terms were arranged in order of decreasing time constants. That is,

$$T_k > T_{k+1} > T_{k+2} \cdots,$$

therefore

$$\omega T_k > \omega T_{k+1} > \omega T_{k+2} \cdots,$$

or

$$\omega T_{k+2} < \omega T_{k+1} < \omega T_k \ll 1. \tag{27}$$

Now, if at the same time it is assumed that $\omega \gg 1/T_{k-1}$, then $\omega T_{k-1} \gg 1$, $(\omega T_{k-1})^2 \gg 1$, and

$$\sqrt{(\omega T_{k-1})^2+1} \approx \sqrt{(\omega T_{k-1})^2} = \omega T_{k-1}.$$

This same type of approximation will hold for all earlier terms of this form in the series, since

$$\cdots T_{k-3} > T_{k-2} > T_{k-1}$$
$$\cdots \omega T_{k-3} > \omega T_{k-2} > \omega T_{k-1} \gg 1.$$

Hence Eq. (26) can be written approximately as

$$|\bar{G}(j\omega)| = \frac{1}{\omega}\,\frac{1}{\omega T_1}\,\omega T_2 \cdots \frac{1}{\omega T_{k-1}} \cdot 1 \cdot 1 \cdots 1 \tag{28}$$

for $1/T_{k-1} \ll \omega \ll 1/T_k$.

This same approximation may be assumed for the interval

$$\frac{1}{T_{k-1}} < \omega < \frac{1}{T_k},$$

as well as for the interval $1/T_{k-1} \ll \omega \ll 1/T_k$ though it cannot be expected to be as good at the extremes of the interval. The magnitude of the errors involved in the approximation is considered below. The result represented by Eq. (28) is the basis for the approximations for $|G(j\omega)|$ given in Column 2 of Table 10·4.

A graphical, and perhaps simpler, basis for passing from the terms of Eq. (25) to those of Eq. (28) is provided if we think of each of the terms of the form $(j\omega T_k + 1)$ as a vector. For then the length of the vector can be regarded as approximated by the length of its longer component if the two components differ markedly in length, as illustrated in Fig. 10·5. On this basis we may skip the algebraic procedure of deriving

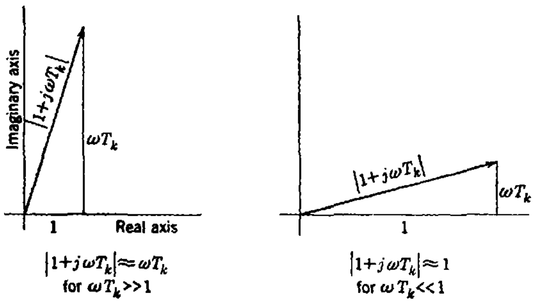


FIG. 10·5.—Vector diagrams illustrating approximation of a vector by its longer component.

the approximation and pass directly from the terms represented in Eq. (25) to those in Eq. (28). For where

$$\omega T_k \gg 1, \qquad |j\omega T_k + 1| \approx \omega T_k;$$

and where

$$\omega T_k \ll 1, \qquad |j\omega T_k + 1| \approx 1.$$

Let us now determine whether or not we can provide a measure of the *errors of approximation* involved in representing a transfer function of any number of factors by means of an asymptotic plot. This is probably done most simply by determining the error involved at any frequency ω_a, due to *any* given factor in the transfer function. For in constructing the asymptotic plot there has been an approximation made for *every* factor in the gain function. Hence the *total* error of approximation will equal the sum of the errors due to each factor present. We shall find that the magnitude of the error due to a given factor, of the form $1/(j\omega T_k + 1)$ or $(j\omega T_k + 1)$, will depend on the ratio of the frequency ω_a to the reciprocal of the time constant of the term considered; i.e., the error will depend on $\omega_a/(1/T_k)$ or $\omega_a T_k$. The sign of the error will depend on whether $(j\omega T_k + 1)$ appears in the numerator or denominator of the transfer function. The error ϵ in decibels is computed by the obvious procedure of determining the difference between the *approximate* expression for the absolute value of the factor and its *exact* value, with each being given in decibels. That is, at a given frequency ω_a, the error corresponding to a term $1/(j\omega T_k + 1)$ will be

$$\epsilon_{\mathrm{db}} = \left|\frac{1}{j\omega T_k + 1}\right|_{\mathrm{db\ approx}} - \left|\frac{1}{j\omega T_k + 1}\right|_{\mathrm{db\ exact}}, \tag{29}$$

$$\epsilon_{\mathrm{db}} = 20 \log \left|\frac{1}{j\omega T_k + 1}\right|_{\mathrm{approx}} - 20 \log \left|\frac{1}{j\omega T_k + 1}\right|_{\mathrm{exact}} \tag{29a}$$

A positive error will indicate that the approximate or asymptotic curve lies above the exact curve; a negative error, that it lies below. The procedure for determining the actual magnitude of errors consists merely in substituting appropriate expressions for the two terms in Eq. (29*a*) and then calculating ϵ_{db} for different values of ωT_k. Two types of cases must be considered separately: (1) those in which $1/T_k < \omega_a$ and (2) those in which $1/T_k > \omega_a$.

1. Let us consider first the cases in which $1/T_k < \omega$, i.e., $1 < \omega T_k$. In such cases,

$$\left|\frac{1}{j\omega T_k + 1}\right|_{\mathrm{exact}} = \frac{1}{\sqrt{(\omega T_k)^2 + 1}},$$

and

$$\left|\frac{1}{j\omega T_k + 1}\right|_{\mathrm{approx}} = \frac{1}{\omega T_k};$$

therefore

$$\epsilon_{\mathrm{db}} = 20 \log \frac{1}{\omega T_k} - 20 \log \frac{1}{\sqrt{(\omega T_k)^2 + 1}} \tag{30}$$

$$= -20 \log \omega T_k + 20 \log \sqrt{(\omega T_k)^2 + 1}. \tag{31}$$

We can now compute the ϵ corresponding to different values of $\dfrac{\omega}{1/T_k}$ or ωT_k. The results are given in Table 10·5.[1] They were obtained by substitution of the specified value of ωT_k in Eq. (31) and use of a table of common logarithms.

2. We may consider now the second type of case, that in which $1/T_k > \omega$, i.e., $1 > \omega T_k$. In such cases,

$$\left|\frac{1}{j\omega T_k + 1}\right|_{\mathrm{exact}} = \frac{1}{\sqrt{(\omega T_k)^2 + 1}}$$

as before, and

$$\left|\frac{1}{j\omega T_k + 1}\right|_{\mathrm{approx}} = 1;$$

therefore

$$\epsilon_{\mathrm{db}} = 20 \log 1 - 20 \log \frac{1}{\sqrt{(\omega T_k)^2 + 1}} \tag{32}$$

$$= 0 + 20 \log \sqrt{(\omega T_k)^2 + 1}. \tag{33}$$

[1] The table previously given (Table 10·3) for the simple case of a transfer function containing a single time-constant factor can be considered a special case of the present table, as will be apparent from a comparison of the entries in the two tables.

The results obtained for various values of ωT_k are shown in the right half of Table 10·5. It will be noted that the table shows a logarithmic symmetry in that the error (except for an altogether negligible difference in two items) is the same for any given value of ωT_k and its reciprocal.

TABLE 10·5.—ERROR OF APPROXIMATION ϵ DUE TO ANY TERM $1/(j\omega T_k + 1)$ IN TRANSFER FUNCTIONS $G(j\omega)$*

$\dfrac{\omega}{\frac{1}{T_k}} = \omega T_k \geqq 1$		$\dfrac{\omega}{\frac{1}{T_k}} = \omega T_k \leqq 1$	
ωT_k	ϵ, db	ωT_k	ϵ, db
1	+3.00	1	+3.00
2	+1.00	$\frac{1}{2}$	+1.00
3	+0.46	$\frac{1}{3}$	+0.45
4	+0.26	$\frac{1}{4}$	+0.25
5	+0.17	$\frac{1}{5}$	+0.17

* When the factor is $(j\omega T_k + 1)$ instead of $1/(j\omega T_k + 1)$, the sign given for the error ϵ is reversed. A positive error indicates that the approximate curve lies above exact curve.

The results of Table 10·5 hold for factors of the form $1/(j\omega T_k + 1)$. By running through the various steps with $(j\omega T_k + 1)$ substituted for its reciprocal, it will be observed that the same results will be obtained, but with a change of sign. This reversal of sign will occur in going from Eq. (30) to (31) and from (32) to (33). Hence for cases with $(j\omega T_k + 1)$ appearing in the numerator instead of denominator, the entries of Table 10·5 can still be used if the sign is reversed.

In applying the results given in Table 10·5 to a given transfer function, one need only determine the values of ωT_k for each time constant T_k appearing in the transfer function, obtain the corresponding error of approximation from the table, and add up all these errors to give the total error. It is apparent from the table that as ωT_k becomes greater than 2 or less than $\frac{1}{2}$, owing to the increasing interval between ω and $1/T_k$, the error due to a given term becomes negligible. The ratio of ω to any given $1/T_k$ can often be obtained most readily from inspection of an asymptotic plot on which the various $1/T$ values have been indicated along the ω-axis as a preparation for drawing in the asymptotic gain curve.

One further point that may require comment is the statement in Step 5 concerning the *slope of the decibel gain curve*. It was stated that each term in the approximate gain function $|\tilde{G}(j\omega)|$ contributes 6 db per octave to the slope of the decibel curve in that interval (see Table 10·4). This result may, perhaps, be inferred from Table 10·4 and the procedure used in determining the slope of the lines in the gain function of the proportional servo. It may be desirable, however, to demonstrate this relationship explicitly for the general case. Let the approximate gain function be given by Eq. (34). We may then determine the gain in decibels at two points ω_a and $2\omega_a$, an octave apart. The *differences* in gain at these two points is 6 db for each term of the form ω or ωT in the gain function $\tilde{G}(j\omega)$, the sign being positive for terms in the numerator and negative for terms in the denominator. Consequently, the slope of the line joining the points will be 6 db per octave for each term, with the same rule for the sign as just mentioned.

$$|\tilde{G}(j\omega)| = \frac{1}{\omega}\frac{1}{\omega T_1}\,\omega T_2 \cdots \frac{1}{\omega T_k} \cdots, \tag{34}$$

$$|\tilde{G}(j\omega)|_{\text{db}} = -20 \log \omega - 20 \log \omega T_1 + 20 \log \omega T_2 \cdots -20 \log \omega T_k \cdots;$$

at $\omega = \omega_a$:

$$\begin{aligned}|\tilde{G}(j\omega)|_{\text{db}} = {}& -20 \log \omega_a - 20 \log \omega_a T_1 + 20 \log \omega_a T_2 - \\ & \cdots -20 \log \omega_a T_k + \cdots;\end{aligned}$$

at $\omega = 2\omega_a$:

$$\begin{aligned}|\tilde{G}(j\omega)|_{\text{db}} = {}& -20 \log 2\omega_a - 20 \log 2\omega_a T_1 + 20 \log 2\omega_a T_2 - \\ & \cdots -20 \log 2\omega_a T_k \cdots \\ = {}& -20 \log \omega_a - 20 \log \omega_a T_1 + 20 \log \omega_a T_2 \cdots - 20 \log \omega_a T_k \cdots \\ & - 20 \log 2 - 20 \log 2 + 20 \log 2 - 20 \log 2.\end{aligned}$$

Difference in gain at ω_a and $2\omega_a$:

$$\begin{aligned}\text{Difference} &= |\tilde{G}(j\omega_{2a})|_{\text{db}} - |\tilde{G}(j\omega_a)|_{\text{db}} \\ &= -20 \log 2 - 20 \log 2 + 20 \log 2 - 20 \log 2 \\ &= -6\text{ db} - 6\text{ db} + 6\text{ db} - 6\text{ db}.\end{aligned}$$

There is thus a 6-db difference for each term, the sign being positive for factors in the numerator and negative for factors in the denominator. It will be noted in Table 10·4 that as we proceed to each succeeding interval, one additional term is added to the gain function, and hence the slope in the new interval changes by 6 db per octave, plus or minus, relative to the slope of the previous interval.

The discussion up to this point has dealt wholly with the procedures for obtaining an approximate plot of *the gain function*. It is necessary to indicate a procedure for plotting *the phase function* as well. The data for the phase function might be obtained, of course, by the straightforward substitution of different values of ω in the transfer function as described in Sec. 10·2. But if the transfer function contains many factors in numerator or denominator or is of high order, the procedure is laborious. The techniques utilized above of writing the transfer function in terms of relational parameters T_1, T_2, . . . , etc., and of regarding separate factors as vectors turn out here too to be useful. Equation (35) gives the general expression for the feedback transfer function used previously. Only the frequency dependent part is considered, since the phase of any constant gain factor will be zero. The component factors are written separately to indicate that each is to be thought of as a separate function. Each is a complex function and may be represented at any given ω as a vector.

The phase of $G(j\omega)$ can be readily obtained from that of the component factors if we apply the rule that the phase of the product of a series of vectors (or complex numbers) is equal to the sum of the phase angles of the separate factors. To obtain the phase angle of the factors containing complex functions in the denominator, we may apply the rule that the phase angle of the quotient of two complex numbers is the phase angle of the numerator term minus the phase angle of the denominator. Or we may regard numerator and denominator as written in exponential form and simply take the difference of the exponents. The phase angle of any term such as $(j\omega T_k + 1)$ will be $\tan^{-1} \omega T_k$ and of $j\omega$ will be $\pi/2$. Hence the expression for the phase angle $\phi(j\omega)$ of $G(j\omega)$ can be written by inspection of Eq. (35) as given by Eq. (36).

$$G(j\omega) = \frac{1}{j\omega}\frac{1}{j\omega T_1 + 1}(j\omega T_2 + 1)\frac{1}{j\omega T_3 + 1} \cdots \frac{1}{j\omega T_k + 1} \cdots, \tag{35}$$

$$\phi(j\omega) = -\frac{\pi}{2} - \tan^{-1} \omega T_1 + \tan^{-1} \omega T_2 - \tan^{-1} \omega T_3 - \cdots - \tan^{-1} \omega T_k - \cdots. \tag{36}$$

Every term in Eq. (35) of the form $(j\omega T_k + 1)$ contributes an angle equal to $\tan^{-1} \omega T_k$ to the phase angle. The sign of this contribution is positive if the term occurs in the numerator and negative if it occurs in the denominator. Equation (36) holds for the entire range of values of ω. However, the contribution of any term in Eq. (36) to the *total* phase angle, for a given value of ω, will depend on the size of the time constant of that term relative to the larger time constants occurring earlier in the series.

To compute the phase angle ϕ for any value of ω, it is necessary only to know the relative magnitudes of the time constants T_1, T_2, $\cdots$, and to substitute the given value of ω in Eq. (36). To illustrate, suppose that the function $G(j\omega)$ contained only the first four factors shown in Eq. (35). Then the phase angle is given by

$$\phi(j\omega) = -\frac{\pi}{2} - \tan^{-1} \omega T_1 + \tan^{-1} \omega T_2 - \tan^{-1} \omega T_3. \tag{37}$$

Suppose it is specified that $T_2 = \frac{1}{2}T_1$ and $T_3 = \frac{1}{5}T_1$. Then

$$\phi(j\omega) = -\frac{\pi}{2} - \tan^{-1} \omega T_1 + \tan^{-1} \frac{\omega T_1}{2} - \tan^{-1} \frac{\omega T_1}{5}. \tag{38}$$

Table 10·6 shows the value of the angles corresponding to the successive terms in Eq. (38), as determined from a table of natural tangents.

TABLE 10·6.—COMPUTATION OF PHASE ANGLES

ω	ωT_1	$\phi(j\omega) = -\frac{\pi}{2} - \tan^{-1}\omega T_1 + \tan^{-1}\frac{\omega T_1}{2} - \tan^{-1}\frac{\omega T_1}{5}$
0	0	$-90° \qquad = -90°$
$\frac{1}{5T_1}$	$\frac{1}{6}$	$-90° - 11° + 6° - 2° = -97°$
$\frac{1}{2T_1}$	$\frac{1}{2}$	$-90° - 27° + 14° - 6° = -109°$
$\frac{1}{T_1}$	1	$-90° - 45° + 27° - 11° = -119°$
$\frac{2}{T_1}$	2	$-90° - 64° + 45° - 22° = -131°$
$\frac{3}{T_1}$	3	$-90° + 72° + 56° - 31° = -137°$
$\frac{5}{T_1}$	5	$-90° - 79° + 78° - 45° = -146°$
etc.		

Reference has already been made (Sec. 10·2) to the formulas and charts developed by Bode for determining the imaginary component from the real component of a network function and conversely. These procedures are relevant here, since the natural logarithm of a transfer function, when expressed in exponential form, gives the logarithmic gain function as the real part and the phase function as the imaginary part as shown by Eq. (4). If the gain function is transformed to a loss or attenuation function, then it will be in the form required for Bode's formulas. It should be noted that the charts are especially designed to permit determination of the imaginary component from the real component when the real component is approximated by a series of straight lines, as is the case in the construction of the asymptotic gain curves discussed above. Bode's book should be consulted for a description of the charts and their use.[1]

In order to complete the method given above for rapid plotting of frequency gain functions, reference should be made to a procedure for dealing with *quadratic factors*. Thus a quadratic factor will occur in the denominator of a feedback transfer function if an elastance as well as inertial and dissipative types of component are involved in a given energy storage unit. This quadratic factor may, of course, then contain

[1] H. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945, Chap. 15, "Graphical Computation of Relations between Real and Imaginary Components of Network Functions, pp. 337–359.

complex roots, a situation that it was not necessary to take into account in dealing with the linear factors so far considered. A mechanical example of a physical unit giving rise to a quadratic factor is the load member of a servo system containing inertia, viscous friction, and mechanical elastance. An electrical analogue is a single mesh containing inductance, resistance, and capacitance in series, as shown in Fig. 10·6.

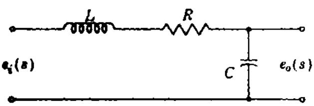


FIG. 10·6.—Electrical circuit giving rise to quadratic factor.

This electrical example will be used here for purposes of illustration, since the dimensional status of combinations of the familiar electrical parameters will be more immediately obvious.

To indicate a method for dealing with such quadratic factors, let us determine the frequency-transfer function of this electrical unit and then determine what procedures can be used in making a rapid plot of its decibel gain curve. The transfer function of the network is readily shown to be

$$P(s) = \frac{e_o(s)}{e_1(s)} = \frac{1}{LCs^2 + RCs + 1}. \tag{39}$$

It is convenient to rewrite this equation in terms of the relational parameters T_q, and ζ, where

$$T_q = \sqrt{LC} \text{ with the dimensions of sec,}$$

and

$$\zeta = \frac{1}{2}\frac{R}{\sqrt{L/C}}, \text{ with no dimensions.} \tag{40}$$

This symbol ζ is exactly equivalent to that defined in Sec. 9·8. The only difference is that it is here expressed in terms of electrical parameters rather than mechanical ones.

From Eq. (40) we obtain

$$LC = T_q^2,$$
$$RC = 2T_q\zeta.$$

Substitution in Eq. (39) gives

$$P(s) = \frac{1}{T_q^2 s^2 + 2\zeta T_q s + 1}. \tag{41}$$

The frequency transfer function is given by

$$P(j\omega) = \frac{1}{(-\omega^2 T_q^2 + 1) + 2\, j\omega T_q \zeta}. \tag{42}$$

Let us now determine the information required for a rapid plot of the decibel gain curve of this transfer function.

For $\omega \ll 1/T_q$, i.e., as $\omega \rightarrow 0$,

$$|P(j\omega)| \approx \frac{1}{1}$$

and

$$|\bar{P}(j\omega)|_{db} = 0. \tag{43}$$

For $\omega \gg 1/T_q$, or $\omega T_q \gg 1$,

$$|P(j\omega)| \approx \frac{1}{|-\omega^2 T_q^2|},$$

and

$$|\bar{P}(j\omega)|_{db} = -20 \log \omega^2 T_q^2, \tag{44}$$

where the symbol $|\bar{P}(j\omega)|_{db}$ has the meaning "approximate gain function."

The slope of this high ω portion of the decibel gain curve can be shown to have a slope of -12 db per octave by the same procedure used previously (comparing $|\bar{P}(j\omega)|_{db}$ for ω_a and $2\omega_a$).

Let us now determine the value of ω at which the low- [Eq. (43)] and high-frequency [Eq. (44)] asymptotes intersect.

$$0 = -20 \log \omega^2 T_q^2;$$
$$\omega^2 T_q^2 = 1;$$

therefore

$$\omega = \frac{1}{T_q}.$$

In case the gain curve has a peak, it will occur at about this value of ω. The height of the gain curve above the 0-db level of the low-frequency asymptote is found by substituting $1/T_q$ for ω in Eq. (42). It is thus found that

$$P(j\omega) = \frac{1}{2j\zeta}, \tag{45}$$

$$|P(j\omega)| = \frac{1}{2\zeta}; \tag{46}$$

$$|P(j\omega)|_{db} = -20 \log 2\zeta. \tag{47}$$

This equation is of interest, since it gives us a direct relation between the value of ζ and the approximate height of the peak in the frequency gain curve. Inspection of the equation shows that for values of ζ less than $\frac{1}{2}$, the decibel gain level is positive, indicating the presence of a peak. For ζ equal to $\frac{1}{2}$, the gain is 0 db; and for ζ greater than $\frac{1}{2}$, the gain level is negative, indicating a tapering off of the gain curve as it approaches the high-frequency asymptote and the absence of a peak.

On the basis of the relations just reviewed, the procedure for plotting the decibel gain function corresponding to a quadratic in the denominator of a transfer function can be summarized as follows: (1) The point corresponding to $\omega = 1/T_q$ is determined on the frequency axis. (2) The gain level at this point, or of the resonant peak if one exists, is given by Eq. (47). (3) The low-frequency asymptote is given by the 0-db axis; the high-frequency asymptote by a line drawn through $\omega = 1/T_q$ with a slope of -12 db per octave. (4) The transition between the asymptotes and the gain at $\omega = 1/T_q$ can be sketched in by hand; or if a more exact determination of the decibel gain in the neighborhood of $\omega = 1/T_q$ is desired, it may be obtained by computing the values of $|P(j\omega)|_{\text{db}}$ corresponding to different values of ω substituted in Eq. (42), or from appropriate charts. (See Vol. 25 of this series.)

For a quadratic factor appearing in the numerator of a transfer function instead of in the denominator, these same rules apply except for a reversal of the sign of the decibel gain level. This will be evident from a review of the development given above, but with the quadratic factor shifted to the numerator.

A final question may be raised concerning the plotting procedure appropriate when additional factors appear in the transfer function, as in Eq. (48).

$$P(s) = \frac{(T_2 s + 1)}{s(T_1 s + 1) \cdots (T_q^2 s^2 + 2\zeta T_q s + 1)}, \tag{48}$$

where $T_1 > T_2 > T_q$.
The answer to this question becomes clear if we regard $P(s)$ factored into parts as given by Eq. (49).

$$P(s) = P_1(s)P_2(s) = \frac{(T_2 s + 1)}{s(T_1 s + 1)} \frac{1}{T_q^2 s^2 + 2\zeta T_q s + 1}, \tag{49}$$

where $P_1(s)$ stands for the factors involving $T_1, T_2, \ldots, T_{q-1}$ and $P_2(s)$ stands for the quadratic factor. It will be evident that $|\bar{P}_1(j\omega)|_{\text{db}}$ can be plotted by the procedures considered earlier and $|\bar{P}_2(j\omega)|_{\text{db}}$ by the method just considered. Consequently, $|\bar{P}(j\omega)|_{\text{db}}$ will be given by the sum of these two gain curves. Hence, if a decibel plot has been made for $P_1(s)$, the quadratic factor is incorporated in this plot by locating on the high ω asymptote of the $P_1(s)$ function the point corresponding to $\omega = 1/T_q$. To the right of this point a line is drawn at a slope of -12 db/octave relative to the asymptote just to the left of this point, and the decibel gain will correspond to that due to both the quadratic and $P_1(s)$ curves. The result of this procedure is illustrated in Fig. 10·7 for the transfer function of Eq. (49).

In the example just considered, the factors present in $P_1(s)$ are all of the linear type. But the same principle of adding decibel gain curves will, of course, hold even when $P_1(s)$ contains quadratic factors in addition to the linear factors.

Decibel vs. Phase-margin Diagram.—The data represented in the decibel vs. log frequency diagram may be plotted in a type of graph known as the phase-margin diagram.[1] The method of plotting consists essentially in plotting decibel gain against phase angle ϕ in rectangular coordinates and then relabeling the abscissa to give the phase angle in terms of *phase margin* rather than *phase angle.* The phase margin ϕ_M in such a diagram is defined as 180° minus the phase lag. Since the phase

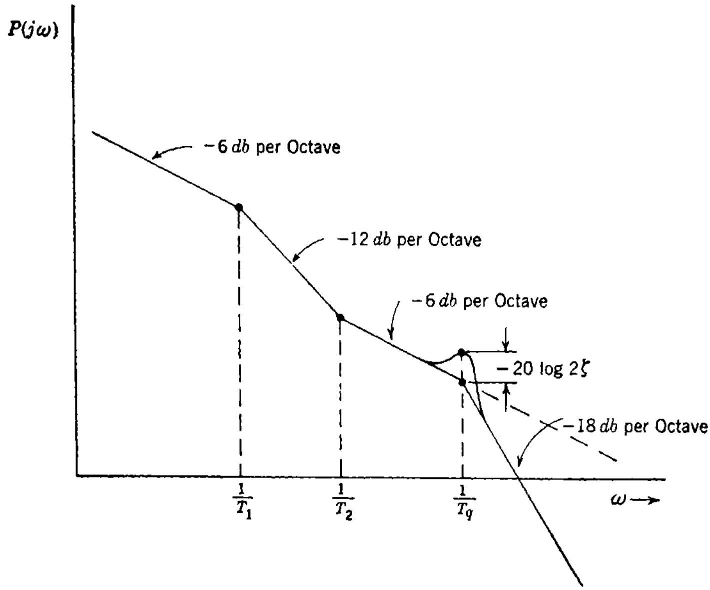


FIG. 10·7.—Asymptotic gain curve for transfer function:

$$P(s) = \frac{(T_2 s + 1)}{s(T_1 s + 1)(T_q^2 s^2 + 2T_q \zeta s + 1)},$$

with actual gain curve in vicinity of $\omega = 1/T_q$.

lag equals the negative of the phase angle,

$$\phi_M = 180° - (-\phi) = 180° + \phi.$$

The above characterization of the phase-margin diagram is illustrated in Fig. 10·8. We may think of the *decibel gain* and *phase* coordinates for various values of ω (e.g., ω_a, ω_b, ω_c, etc.) as plotted in relation to the coordinate axes drawn in solid lines. Then at a phase angle of −180° a new vertical axis is drawn (shown as a broken line) and used as the reference for phase-margin measurements. It is obvious from inspection of the scales along the abscissa that the phase margin will equal 180° plus the phase angle. In plotting frequency-response data in such a diagram,

[1] A. Sobczyk, RL Report No. 811, 1946; and D. P. Campbell, Nichols lecture, *loc. cit.* See also Vol. 25, Radiation Laboratory Series.

it will, of course, not be necessary to make use of the phase reference axis, since the phase margin can easily be obtained from the phase angle by adding 180°. It is of interest to note the likeness of the phase-margin diagram to the transfer locus in that both involve essentially a plot of gain against phase. The two differ in that the one is in rectangular coordinates whereas the other is in polar coordinates and in the use of the decibel scale for representing gain in the phase-margin diagram.

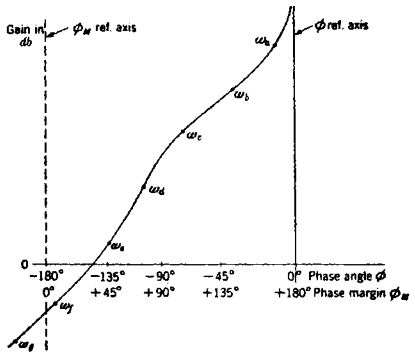


FIG. 10·8.—Phase-margin diagram.

10·4. The Interpretation of Frequency Diagrams.—The previous section summarizes different ways in which the frequency-response curves can be plotted. In the present section will be considered the ways in which these graphs may be interpreted to yield predictions of the performance of the system for various excitation conditions. Such estimates are not limited to input signals of periodic character. The specific question that this section will attempt to answer is this—from any given type of graph, what predictions can be made concerning the performance of the system represented? The discussion will be concerned principally with the feedback transfer function $KG(j\omega)$.

The Output Transfer Function.—In order to develop criteria for interpretation of the transfer locus, it is helpful to consider first the output transfer function and its relation to the transient response of the system. From the correlations thus established, it will be possible to proceed to the transfer locus, which represents the feedback transfer function, and establish techniques for relating it to the transient response. The amplitude and phase curves of $\theta_o(j\omega)/\theta_i(j\omega)$ may simply be plotted as two separate curves as shown in Fig. 10·9. The logarithmic plot need not be used.

A preliminary consideration of the response curves of an ideal system (one in which the output follows the input immediately) indicates the bearing of the frequency-response curves on the transient response. In an ideal system, the amplitude function will be equal to 1 and the phase response equal to 0 for the entire frequency range from 0 to ∞.[1] A system with such response curves would give perfect following, as is evident from Fourier integral concepts. Thus, any nonperiodic input disturbance can be represented by a specific continuous frequency spectrum extending from $-\infty$ to $+\infty$. If each of the component frequencies is transmitted by the servo system without change in amplitude or phase (as implied by a constant amplitude response of 1), then each of the input frequencies will reappear at the output in its original magnitude and phase. The recombination of these component frequencies by means of a Fourier synthesis will, therefore, reproduce the original input disturbance.

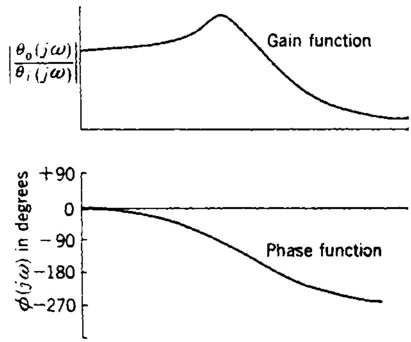


FIG. 10·9.—Amplitude and phase response curves of output transfer function $\theta_0(j\omega)/\theta_i(j\omega)$.

No physical system is, of course, capable of showing this ideal response. The same concepts are, however, applicable. The arbitrary input signal, expressed as a function of time, can again be represented by a Fourier spectrum. The frequency-response curves of the total system will indicate the extent to which each frequency component in the signal is transmitted by the system, i.e., the change that it undergoes in amplitude and phase. More precisely, the frequency-amplitude curve representing the input signal multiplied by the amplitude transfer curve of the system will give the Fourier spectrum of the output signal. The phase function of the output might be obtained similarly by adding the phase function of the system to the phase function representing the input time signal. The output signal expressed as a time function, or transient response, may again be obtained by the Fourier synthesis of the component frequencies. These considerations indicate that it is reasonable to expect a correlation between the properties of the transient-response and the frequency-response curves for the total system.

By way of a practical though approximate index of the transient response, Hall has reported the following correlation, based on a comparison of frequency-response curves and transient response for the same systems. The presence of peaks in the amplitude response is generally associated with the occurrence of complex roots of the characteristic

[1] In a minimum phase system, the specification of a constant amplitude of 1 would be sufficient to indicate a constant phase of 0.

equation. The height of the peak, relative to the flat, low-frequency part of the curve, is an index of the real part of the root, tending to increase as the magnitude of the real part decreases. The angular frequency at which the peak occurs is an index of the imaginary part of the root, tending to increase as the magnitude of the imaginary part increases. Hence the frequency of the peak will be an index of the frequency of the oscillatory component of the transient response, and the height an index of its damping.[1] In numerical terms, Hall states that if the height of the peak (relative to the response at $\omega = 0$) is limited to $1\frac{1}{3}$, then the damping ratio ζ will lie between 0.5 and 0.8, and the angular frequency of the peak will equal the frequency of oscillation to within about 20 per cent. A damping ratio of about 0.8 has been proposed as a useful practical standard in the design of many systems, for it provides for a quick transient response with relatively little overshoot.[2] Limitation of the resonant peak to $1\frac{1}{3}$ thus provides a criterion that can be used in adjusting the parameters of the system.

A possible qualitative basis for this type of correlation may exist in the relations derived in Sec. 10·3 between the damping ratio ζ and the form of the gain curve of a transfer function consisting of a quadratic factor. Let us take as an example a proportional servo system. Its feedback transfer function is

$$KG(s) = \frac{\theta_o(s)}{E(s)} = \frac{k_0}{Js^2 + fs}.$$

The output transfer function is found from this to be

$$\begin{aligned}\frac{\theta_o(s)}{\theta_i(s)} &= \frac{KG(s)}{1 + KG(s)} = \frac{k_0}{Js^2 + fs + k_0} \\ &= \frac{1}{\frac{J}{k_o}s^2 + \frac{f}{k_o}s + 1}.\end{aligned} \tag{50}$$

This equation can be written in terms of the relational parameters ζ and T_q considered previously by setting

$$\zeta = \frac{f}{2\sqrt{k_oJ}} \qquad \text{(nondimensional)}$$

and

$$T_q = \sqrt{\frac{J}{k_o}} \qquad \text{(with dimensions of sec).}$$

[1] It may be helpful to refer back at this point to Sec. 9·8 on the relation of the transient response to the real and imaginary parts of the complex roots of a quadratic characteristic equation.

[2] H. Harris, "The Analysis and Design of Servomechanisms," NDRC Report, 1942, p. 11, points out, however, that although this criterion is particularly useful in the design of regulators, additional factors must be considered in the case of servomechanisms.

These parameters are exactly equivalent to those considered previously, the only difference being that here they are defined in terms of mechanical rather than electrical physical parameters. Equation (50) written in terms of the new parameters becomes

$$\frac{\theta_o(s)}{\theta_i(s)} = \frac{1}{T_\varphi^2 s^2 + 2\zeta T_\varphi s + 1}.$$

It will be noted that this equation is identical with that given previously [Eq. (41)] in discussing the nature of the gain function plot of a quadratic factor. Hence the conclusions derived then apply here, and the value of the gain function at $\omega = 1/T_\varphi$, the approximate location of the peak in the $\theta_o(j\omega)/\theta_i(j\omega)$ function, will be given by

$$\left|\frac{\theta_o(j\omega)}{\theta_i(j\omega)}\right| = \frac{1}{2\zeta}.$$

We thus find an inverse relationship between the damping ratio ζ and the height of the peak that suggests the possibility of utilizing the height of the peak as an index of the damping and stability of the system.

Transfer Locus.—Although graphs of the output transfer function are useful for estimating the *performance* of a system, they are not so convenient as a basis for *design* as are the transfer loci. Certain of the measures of performance given for the $\theta_o(j\omega)/\theta_i(j\omega)$ curves are therefore used in developing criteria that may be applied to the $\theta_o(j\omega)/E(j\omega)$ loci. In the design and adjustment of servo systems the goal of the designer is to obtain a system that will be stable and will meet certain specifications of accuracy. The following material on the interpretation of transfer loci will therefore be considered in relation to these two topics.

Stability.—The criterion of servo-system stability utilized in the frequency approach is based on Nyquist's theoretical analysis of regeneration in feedback amplifiers.[1] A servo system may be considered as dynamically analogous to a negative feedback amplifier.[2] The criterion of stability developed for the latter may, therefore, be carried over and applied to servo systems.

To apply this criterion it is necessary to determine whether or not the transfer locus of $KG(j\omega)$, corresponding to values of ω from $-\infty$ to $+\infty$, encircles the critical point $-1 + j0$.[3] If the locus encircles the

[1] H. Nyquist, "Regeneration Theory," *Bell System Tech. Jour.*, **11**, 125–147, January 1932.

[2] See A. C. Hall, *Analysis and Synthesis of Linear Servomechanisms*, Technology Press, Massachusetts Institute of Technology, 1943, pp. 34*ff*.

[3] Nyquist's criterion originally formulated (*op. cit.*) for feedback amplifiers was given in terms of the relation of the polar plot of the feedback function $\mu\beta$ to the point $+1 + j0$. Since in terms of the symbols used here $\mu = KG(j\omega)$ and $\beta = -1$, the servo

critical point, the system represented is unstable; otherwise it is stable. Figure 10·10 shows examples of loci of the two types. A and B are stable loci; C and D, unstable loci. The loci in the figure have been drawn only for positive values of ω. The part of any locus corresponding to negative values of ω will be the mirror image in the real axis of the part drawn, since the value of the vector $KG(j\omega_a)$, for any specific frequency ω_a, will be the conjugate of the vector $KG(-j\omega_a)$. The locus corresponding to negative values of ω may therefore be sketched in readily.

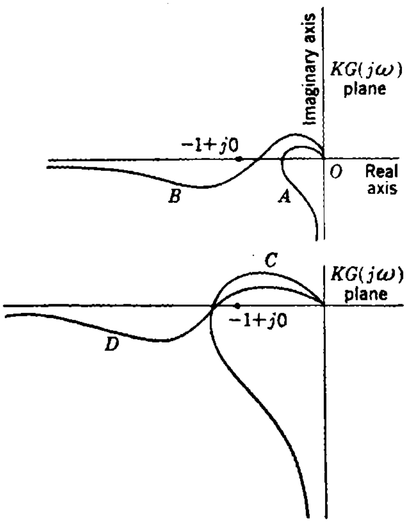


FIG. 10·10.—Examples of transfer loci representing stable (A and B) and unstable (C and D) servo systems.

In some cases, it may be difficult to decide whether or not the critical point is enclosed. For such cases, the following supplementary rule proposed by Nyquist is of help. Assume that a straight line is drawn between the critical point $-1+j0$ to any point on the locus of the function $KG(j\omega)$. As the frequency varies from $+\infty$ to $-\infty$, let the end of the line on the locus trace out the locus, while the end of the line at the critical point remains fixed. If the final angle (net angle) of swing of the straight line is zero as the frequency varies from $+\infty$ to $-\infty$, then the system is stable; if it is different from zero, the system is unstable.[1] The unstable systems of Fig. 10·10 may be used as examples in trying out this method. In applying the rule, a pencil laid over the graph may be used to represent the straight line.

Special difficulties of interpretation may arise in the case of loci that extend to ∞ along the negative real axis such as that of B of Fig. 10·10. In applying the Nyquist criterion to loci of this type, the part of the locus approaching $-\infty$ should be regarded as connected by a circle of

---

feedback transfer function would be represented as $-KG(j\omega)$ and its relation to the point $+1\ +j0$ determined. The stability criterion can, however, be just as well formulated in terms of the relation of $+KG(j\omega)$ to the $-1\ +j0$ point, which is equivalent to the relation of $-\mu\beta$ to the $+1\ +j0$ point. The tendency has been for the Nyquist criterion to be used in this form. See, for example, Hall, *op. cit.*, p. 36; and H. W. Bode, *Bell System Tech. Jour.*, **19**, 421–454, July 1940. For a discussion of Nyquist's criterion in relation to complex function theory, see H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945, pp. 137–169.

[1] Taken, with slight modifications, from Hall, *op. cit.*, pp. 35*ff*.

infinite radius as shown in Fig. 10·11. The appropriateness of this procedure becomes clear when it is noted that vectors corresponding to

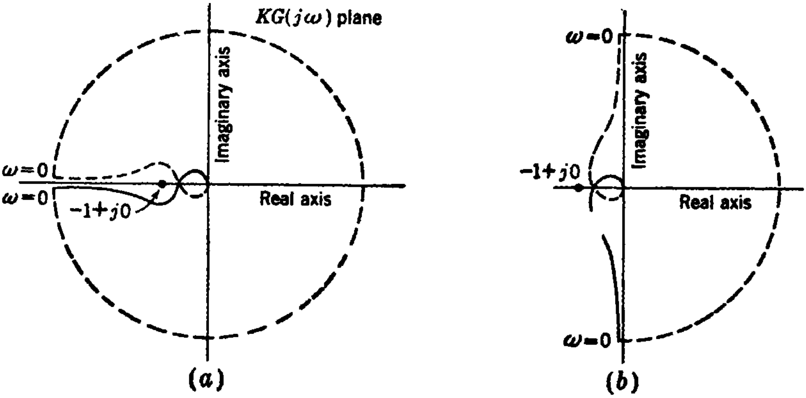


FIG. 10·11.—Use of infinite circle in application of Nyquist's criterion for stability.

positive frequencies (solid line) are approaching a phase angle of $-180°$ as $\omega \to 0$, whereas vectors corresponding to negative frequencies are approaching a phase angle of $+180°$ as $\omega \to 0$. These phase angles are therefore to be regarded as 360° apart.[1]

Loci such as E of Fig. 10·12 have given rise to a distinction between "absolute" stability and "conditional" stability.[2] Curve E_1 represents a "conditionally" stable system. As drawn here the locus does not enclose the critical point. The corresponding system is therefore stable. If, however, the gain of the system is *increased*, to give Curve E_2 as its locus, or the gain is *decreased*, to give Curve E_3 as its locus, then the system is no longer stable. The term "Nyquist stability" has also been used as a synonym for "conditional" stability.

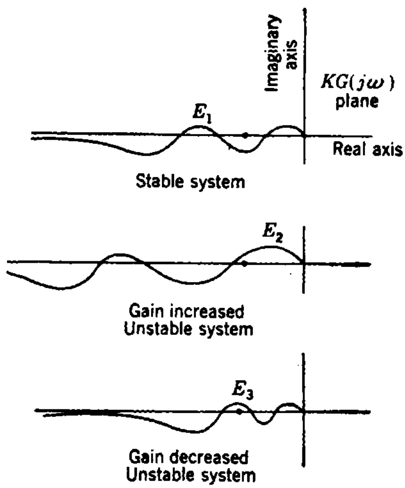


FIG. 10·12.—Effect of variation in gain on transfer locus of system with conditional or "Nyquist stability."

[1] See Hall, *op. cit.*, p. 40; Brown and Hall, *op. cit.*, p. 28; and McColl, *Servomechanisms*, Van Nostrand, New York, 1945, pp. 28*ff.*

[2] See H. W. Bode, *Bell System Tech. Jour.*, **19**, 421–454, July 1940; and *Network Analysis and Feedback Amplifier Design*, pp. 162–164.

Nyquist stability is not, in general, satisfactory in a servo system, since changes in any of the factors that influence the over-all gain (such as loss of tube gain with age or increase of tube gain from zero as power is first applied to the amplifier)[1] may result in an unstable system.

In the adjustment of a servo system in order to provide stability, it is evident that a system which barely passed the test of stability (e.g., curve S of Fig. 10·13) might not be satisfactory. Slight uncontrollable changes in the parameters of the system, such as occur in aging, might produce sufficient change in the locus of the system to move it over to the other side of the critical stability point (curve U of Fig. 10·9). In the adjustment of such systems, it has, therefore, been found desirable to provide *margins of safety*. Bode[2] has discussed the theoretical considerations involved in the selection of such margins in the design of feedback amplifiers and introduces the concepts of phase and gain margins.[3] Ferrell[4] in discussing the application of Bode's treatment to servo systems states that a phase margin of between 40° and 60° and a gain margin of 10 to 20 db constitute good design practice. These criteria mean that at the frequency at which the gain has fallen to 0 db, the phase margin should be not less than 40° to 60° (i.e., the phase lag should not exceed 120° to 140°); when the phase lag has reached 180°, the gain in decibels should have fallen to between -10 and -20 db. Hence in the interpretation of loci, one must determine not only whether or not the system represented is stable but also whether or not adequate or standard margins of safety have been provided.

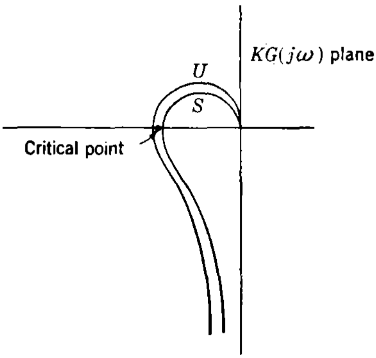


FIG. 10·13.—Diagram illustrating the need for safety margins in the adjustment of stability.

It should be noted that the principles given above relative to system

[1] This type of situation may lead to damage to the system due to excessive oscillations that may occur for low-gain values of the amplifier before the system becomes stable at the higher-gain values associated with steady-state temperature of the tube cathodes.

[2] H. W. Bode, "Relations between Attenuation and Phase in Feedback Amplifier Design," *Bell System Tech. Jour.*, **19**, 433–436, July 1940.

[3] The term phase margin has already been introduced, it will be recalled, in Sec. 10·3, in the description given of a method for plotting decibel-phase-margin diagrams.

[4] Ferrell, "The Servo Problem as a Transmission Problem," *Proc. IRE*, **33**, 763-767, November 1945.

stability apply equally well to *regulators*,[1] that is to say, to systems designed to maintain some property *constant* on the basis of a feedback loop. A regulator can be regarded as a servo system in which the input function $\theta_i(t)$ is constant. A block diagram of a d-c voltage regulator regarded from this point of view is given in Fig. 10·14. Comparison of this diagram with one used to represent a servo system (e.g., Fig. 9·8) shows that there are no essential differences in the structure of the two systems. The differences consist rather in the nature of the disturbances that tend to produce variations in the output quantity $e_o(t)$. In the voltage regulator, the origin of such variations is in the d-c voltage source applied to the regulator tube and in the parameters of the load circuit

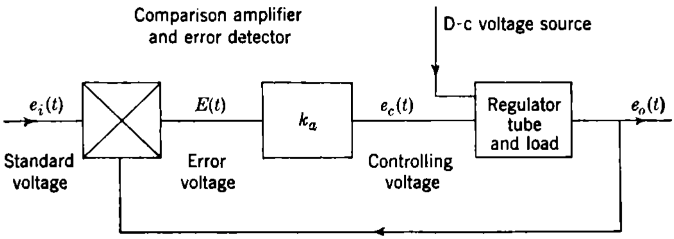


FIG. 10·14.—Block diagram of voltage regulator.

rather than in the input forcing function $e_i(t)$. Since the latter is purposely kept constant, the output response $e_o(t)$ tends to return to this same value following disturbances in the voltage source or in the load.

Accuracy.—As shown above, the transfer locus provides a simple and precise basis for inferences concerning the stability of a system. It is much less satisfactory as a basis for estimates of accuracy. These limitations are not, however, inherent in the frequency approach. As we shall see, a shift to the decibel log frequency methods of representation permits certain conventional estimates of servo error to be made with considerable precision. Before going on to consider these procedures, let us review the chief relations that have been established between transfer loci and measures of servo error. Hall[2] has pointed out an interesting series of relations between conditions of *zero steady-state* error and the shape of the locus as the angular frequency ω approaches zero. These correlations are represented in Fig. 10·15.

In systems with zero displacement error, the locus approaches ∞ along the negative imaginary axis as ω approaches zero. This relationship is shown in Fig. 10·15*a*. The specification of a zero displacement error means that the steady-state value of the error will be zero if the input

[1] See Sec. 8·3 for a definition of regulators.

[2] Hall, *op. cit.*, pp. 38–41.

$\theta_i(t)$ consists of a *fixed displacement*, such as occurs in the case of a step-function input.

In systems with zero velocity error, the locus of the transfer function approaches ∞ along the negative real axis as ω approaches zero. This relation is shown in Fig. 10·15*b*. The specification of zero velocity error means that the steady-state error will be zero if the input $\theta_i(t)$ consists of a fixed velocity, such as occurs for a step-velocity input.

In systems with zero acceleration error, the locus approaches ∞ along the positive imaginary axis as ω approaches zero. This is shown in Fig. 10·15*c*. Specification of zero acceleration error means that the steady-state error will be zero for a fixed input acceleration.

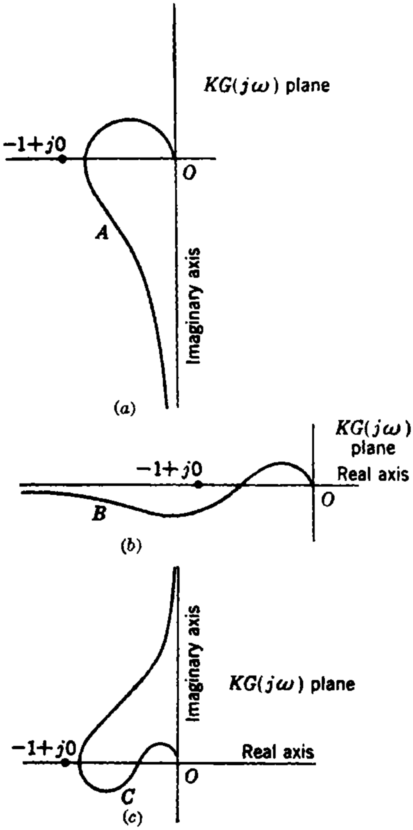


FIG. 10·15.—Transfer locus diagrams representing systems with zero steady-state errors. (*a*) zero displacement error; (*b*) zero velocity error; and (*c*) zero acceleration error.

It is evident that as the *order* of the zero steady-state error increases, the axis along which the locus approaches ∞ (when $\omega \to 0$) shifts progressively in a clockwise direction. We might, on this basis, readily lay down the requirements for zero steady-state errors of still higher order. But systems are seldom required to meet such higher-order specifications, since the input functions that occur commonly do not tend to show constancy in the higher-order derivatives of the displacement. Inputs approximating constant velocities (the first derivative of the displacement) are probably most common.

The correlations just given are useful in comparing a proposed servosystem with performance specifications. By noting whether or not the low-frequency end of a given locus approaches ∞ along the correct axis, one may determine the nature of corrections needed in the system to provide a locus of the required type.

What is the reason for these correlations? To understand them, two facts must be considered: (1) the properties that specification of a given steady-state error as zero imposes on the transfer function $KG(s)$ and (2) the nature of the behavior of the corresponding $KG(j\omega)$ function as $\omega \rightarrow 0$.

The first relationship is made concrete by considering a specific system, such as the proportional servo of previous illustrations. Its error function is

$$E(s) = \frac{Js^2 + fs}{Js^2 + fs + k_0}\,\theta_i(s). \tag{51a}$$

Suppose the system is required to have a zero displacement error. This implies that when $\theta_i(t)$ is a step function, there must be no constant terms in the solution $E(t)$. Substituting $1/s$ for $\theta_i(s)$,

$$E(s) = \frac{Js^2 + fs}{Js^2 + fs + k_0}\,\frac{1}{s}. \tag{51b}$$

If there were no factor s in the numerator to cancel the s introduced into the denominator by $\theta_i(s)$, then, upon application of the inverse Laplace transformation, there would be a constant term in the equation for $E(t)$, corresponding to the factor s in the denominator of $E(s)$. This follows from transform pair (a) of Table 9·1a, and (e) of Table 9·1b. The constant term would constitute the steady-state error, since it would be unchanged as $t \rightarrow \infty$. Hence, specification of a zero steady-state error requires absence of s as a separate factor in the denominator of $E(s)$.

But since an s is introduced into the denominator by $\theta_i(s)$, a factor s must be present in the numerator of the transfer function $E(s)/\theta_i(s)$ to cancel it. This requirement is met in the case of Eq. (51b) of our example, since an s factor is available in the numerator to cancel that of the denominator. But we know that the numerator of the error transfer function $E(s)/\theta_i(s)$ is identical with the denominator of the feedback-transfer function $KG(s)$, as will be evident from inspection of Eqs. (9.22) and (9.47) showing derivation of $E(s)/\theta_i(s)$ from $KG(s)$. Therefore to cancel a first-order pole introduced by $\theta_i(s)$ into $E(s)$, there must be a first-order pole in $KG(s)$; to cancel a second-order pole, there must be a second-order pole in $KG(s)$; and so on.

Hence, specification of zero displacement error means that $KG(s)$ is of the form

$$KG_1(s) = \frac{(s + a_1)(s + a_2)(s + a_3) \cdots}{s(s + b_1)(s + b_2)(s + b_3) \cdots} \tag{52a}$$

Specification of zero velocity error implies $KG(s)$ is of form

$$KG_2(s) = \frac{(s + a_1)(s + a_2)(s + a_3) \cdots}{s^2(s + b_1)(s + b_2)(s + b_3) \cdots}, \tag{52b}$$

and so on

We are now ready to consider the nature of the transfer loci corresponding to Eqs. (52a) and (52b). The frequency-transfer function corresponding to Eq. (52a) will be

$$KG_1(j\omega) = \frac{(j\omega + a_1)(j\omega + a_2)(j\omega + a_3) \cdots}{j\omega(j\omega + b_1)(j\omega + b_2)(j\omega + b_3) \cdots}. \tag{52c}$$

What happens to the various factors as ω approaches zero? Each of the factors of the form $(j\omega + a)$ and $(j\omega + b)$ are complex numbers. The terms represented by $a_1, a_2, a_3, \ldots$, and $b_1, b_2, b_3, \ldots$, may be real or complex. As $\omega \to 0$, $j\omega$ becomes negligible relative to the other part of the factor; therefore the $j\omega$ term inside each parenthesis can be neglected, and each factor can be approximated by its root alone.

Therefore

$$\lim_{\omega \to 0} KG_1(j\omega) \approx \frac{a_1 a_2 a_3 \cdots}{j\omega(b_1 b_2 b_3) \cdots}. \tag{53a}$$

The products $a_1a_2a_3a_4 \ldots$ and $b_1b_2b_3b_4 \ldots$ will be real, since if any of the single roots is complex, it will be paired with a conjugate root and the product of the two will be real. The right-hand side of Eq. (53a) will therefore be an imaginary number that approaches $-j\infty$ as ω approaches zero. Hence the function $KG_1(j\omega)$ will approach infinity along the $-j$ axis as ω approaches zero.

A similar line of reasoning indicates that if a zero velocity error is prescribed, then $\theta_i(s)$ is represented by N/s^2, and $KG_2(s)$ must contain an s^2 term in its denominator, as shown by Eq. (52b). The corresponding frequency-transfer function is

$$\begin{aligned} KG_2(j\omega) &= \frac{(j\omega + a_1)(j\omega + a_2) \cdots (j\omega + a_n)}{j^2\omega^2(j\omega + b_1)(j\omega + b_2) \cdots (j\omega + b_n)} \\ &= \frac{(j\omega + a_1)(j\omega + a_2) \cdots (j\omega + a_n)}{-\omega^2(j\omega + b_1)(j\omega + b_2) \cdots (j\omega + b_n)}; \end{aligned}$$

therefore

$$\lim_{\omega \to 0} KG_2(j\omega) \approx \frac{a_1 a_2 a_3 \cdots a_n}{-\omega^2 b_1 b_2 b_3 \cdots b_n}. \tag{53b}$$

This is a real number which will lie on the negative real axis and will approach $-\infty$ as ω approaches zero. Hence the function $KG_2(j\omega)$ will approach ∞ along the negative real axis as ω approaches zero.

We might proceed by a similar line of reasoning to sketch the behavior, in the low-frequency region, of the transfer loci corresponding to zero steady-state errors of still higher orders. The examples given above, however, should be sufficient. It may be of interest, before leaving this topic, to note the characteristics of the transfer function $KG(j\omega)$ which determine the behavior of the locus at its *high*-frequency end. As Hall points out,[1] the behavior of the locus in this region depends on the *order* of the transfer function. The order of the transfer function may be defined, on the basis of the following equation, as the difference $(q - p)$.

$$\underset{\omega \to \infty}{G(j\omega)} = \frac{(j\omega)^p}{(j\omega)^q}. \tag{54}$$

This equation shows the form taken by the frequency-transfer function representing any given system, [Eq. (52*c*)] for large values of ω. As $\omega \to \infty$, the transfer locus will approach zero from a direction determined by the order $(q - p)$. Thus, if $(q - p) = 2$, Eq. (54) becomes

$$G(j\omega)]_{\omega \to \infty} = \frac{(j\omega)^{q-2}}{(j\omega)^q} = \frac{1}{(j\omega)^2} = -\frac{1}{\omega^2};$$

i.e., $G(j\omega)]_{\omega \to \infty} \to -1/\omega^2$. For large values of ω, the transfer locus approaches zero along the negative real axis.

If $(q - p)$ is assumed to equal 3, Eq. (54) becomes

$$G(j\omega)]_{\omega \to \infty} = \frac{1}{(j\omega)^3} = -\frac{1}{j\omega^3} = +j\frac{1}{\omega}.$$

For large values of ω, $G(j\omega)$ approaches 0 along the + imaginary axis.

If $(q - p)$ is assumed to equal 4, Eq. (54) becomes

$$G(j\omega)]_{\omega \to \infty} = \frac{1}{(j\omega)^4} = +\frac{1}{\omega^4}.$$

For large values of ω, $G(j\omega)$ approaches 0 along the + real axis. The shape of the locus at the high-frequency end may thus be readily determined from the order of the transfer function.

As Hall points out, the interpretation thus indicated for different regions of the locus permits the locus of a given system to be sketched rapidly with a minimum of computation. The shape of the locus at the high-frequency end is indicated by the *order* of the transfer function, which is determined by the number of energy-storage devices in the system. The shape at the low-frequency end is indicated by the specifications regarding the required zero steady-state error. Finally, the required relation of the locus to the critical point $(-1 + j0)$ is indicated by Nyquist's stability criterion.

In its ability to provide estimates of the *transient error,* the frequency approach, as represented by curves of the output transfer function or by transfer locus plots, is again not quite satisfactory. The decibel-log ω plots here, too, turn out to be somewhat more useful. Some general correlations have, however, been reported. It will be recalled from

[1] Hall, *op. cit.*, pp. 41*ff*.

Sec. 9·8 that the magnitude of real roots of the characteristic equation of the error function, if negative, determines the rate of decay of exponential components of the error response and that the magnitude of the real part of complex roots, if negative, determines the rate of decay of any oscillatory component. In both cases, the greater the absolute magnitude (of the real root, or the real part of a complex root) the more rapid is the decay, that is, the shorter is the transient. Hall[1] reports, from his comparison of transient and frequency-response curves, that the height of the peak of the amplitude curve of the output transfer function can be used as an index of the size of real roots and the real parts of complex roots. He states that in order for these to be large, "the peaks in the amplitude response function must be limited in magnitude and occur at large frequencies."

The specification concerning the peaks in the amplitude function can be carried over to the transfer locus. A simple geometrical procedure to be described in Sec. 10·5 permits one to determine, on the locus, the frequency at which such a peak in the amplitude response [of $\theta_o(j\omega)/\theta_i(j\omega)$] will occur. The reported association between high frequency of the peak and short duration of the transient, although not particularly useful in making possible quantitative estimates of the duration of the transient, will be found of value in the design of corrective devices that are intended to bring about an increase in frequency of the peak through a counter-clockwise rotation of the transfer locus.

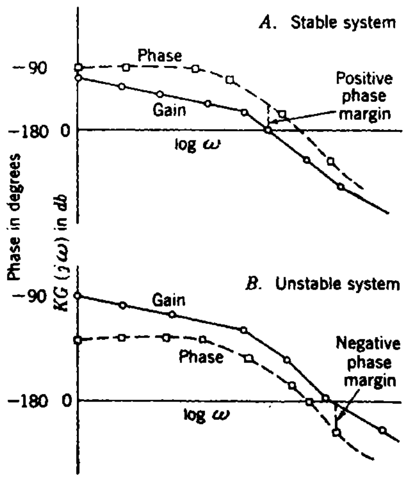


FIG. 10·16.—Decibel vs. log frequency plots illustrating relations of gain and phase curves in stable and unstable systems.

Decibel-log Frequency Diagrams. The data represented in the feedback transfer function can be plotted as a log-log plot in the way already described, instead of as a polar plot of gain against phase. This type of plot, although based on the same data, represents it in a different manner and is more useful for some purposes. As was to be expected, we find some of the same criteria discussed in the section above reformulated in terms of the new curves. The interpretation of these curves is considered here, too, in relation to the topics of *stability* and *accuracy*.

[1] Hall, *op. cit.* p. 17.

Stability.—Figure 10·16 shows decibel-log ω plots of stable and unstable systems such as are represented in the polar plots of Fig. 10·10: The upper curves (A) correspond to a stable system; and the lower curves (B) to an unstable system. Nyquist's criterion is again used but formulated now in terms of the new method of representation.

It will be sufficient to formulate the criterion for an absolutely stable system, such as those represented by Curves A and B of Fig. 10·10.[1] In terms of *Nyquist's criterion,* the essential requirement for the system to be stable[2] is that the gain be less than one by the time, with increasing frequency, the phase lag reaches 180°. In our db-log ω diagram of Fig. 10·16 the gain and phase curves are plotted relative to the same axis. The horizontal axis, for the gain curve, indicates 0 db. Regions below the axis indicate gains less than one. For the phase curve, the horizontal axis represents a phase lag of 180°, and regions below the axis, phase lags greater than 180°. If the gain and phase curves cross the axis at the same point (as in Fig. 10·17), it would mean that at the frequency at which the phase lag reached 180°, the gain was exactly one. This condition would correspond to a transfer locus crossing the negative real axis at the critical stability point $-1 + j0$. The system would therefore be unstable. In order for the system to be stable, the gain must be less than one at this value of the phase. Consequently, Nyquist's criterion may be formulated as follows. *In an absolutely stable system, the gain curve must cross the* 0-db *axis at a lower frequency than that at which the phase curve crosses the* $-180°$ *phase axis.* In the diagram, the phase crossover point must lie to the right of the gain crossover point, as in Fig. 10·16a.

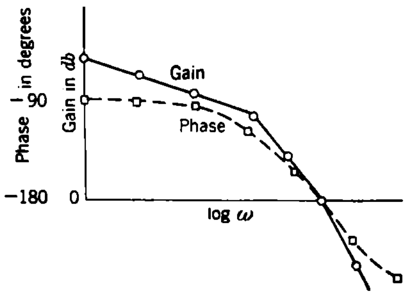


FIG. 10·17.—Decibel vs. log frequency diagram of a system just at the boundary line between stability and instability. Gain and phase margins are zero.

Stability Margins.—The rule just stated may now be reformulated to provide the margins of safety considered in connection with transfer loci. Ferrell, it will be recalled, proposed as good design practice a *phase margin* of 40° to 60° at the gain crossover point and a *gain margin* of 10 to 20 db at the phase crossover point. The definitions given of gain and phase margins are illustrated in Fig. 10·2.

[1] The reader, if he so desires, should have no difficulty in formulating a similar rule sufficiently general to include conditionally stable systems, through inspection of Fig. 10·12 and equivalent curves drawn for a decibel-log ω plot.

[2] See, for example, H. W. Bode, *Bell System Tech. Jour.*, **19**, 432, July 1940.

Accuracy.—Ferrell[1] has pointed out that the asymptotic decibel-log ω gain curve may be used to provide measures of system accuracy. His method is important, since it makes up for what was previously a deficiency in the frequency approach. Ferrell's derivation of the method is given only for the proportional servo system, characterized by a second-order transfer function. In the present discussion, a brief summary of Ferrell's account will be given first, retaining the symbols used by him to represent measures of servo-system error. The relation of his parameters to the time-constant set (k_v, T, ω_n) described in Sec. 9·8 will be pointed out, and finally a proof will be given of the applicability of Ferrell's method to determination of the steady-state velocity error of systems with higher-order transfer functions.

Ferrell's method may be understood by reference to the asymptotic plot of Fig. 10·3, which gives the curve for the loop gain μ of a proportional servo system. The equation given for μ, the loop gain of the system [equivalent to our feedback transfer function $KG(s)$], is[2]

$$\mu = \frac{S_m\mu_m}{S + pR + p^2J}, \tag{55}$$

where S_m = conversion constant of a potentiometer that converts input mechanical displacement to volts,

μ_m = motor conversion constant, in torque units per volt,

S = elastance or stiffness of the motor load,

R = resistance of the load, regarded as including both the motor's internal resistance and the viscous friction of the load,

J = inertia of the load,

p = differential operator d/dt.

The equation given for the error $\Delta\theta$ is

$$\Delta\theta = \frac{\theta}{\mu}. \tag{56}$$

Here θ is regarded as representing either the input θ_1 or the output θ_2, an approximation considered justifiable by Ferrell when the loop gain μ is assumed to be very large. On this basis, this equation can be regarded as merely equivalent to the definition of the loop gain, stating (after interchanging μ and $\Delta\theta$) that the loop gain equals the output divided by the error.

[1] E. B. Ferrell, "The Servo Problem as a Transmission Problem," Bell Telephone Laboratories Report No. MM45-180-6, Jan. 27, 1945; also *Proc. IRE*, **33**, 763–767, November 1945.

[2] This equation corresponds to Eq. (9·21), Sec. 9·3. The only difference in the systems represented is that Ferrell regards the output or load member as including elastance, represented by the parameter S, in addition to inertia and dissipative parameters.

If, now, it is assumed that the load elastance $S = 0$ and the terms ω_0 and ω_1 are substituted in Eq. (55), then it can be rewritten

$$\mu = \frac{\omega_0\omega_1}{p(p + \omega_1)}, \tag{57}$$

where

$$\omega_0 = \frac{S_m\mu_m}{R},$$

and

$$\omega_1 = \frac{R}{J}.$$

If this value of μ is substituted in Eq. (56), we obtain

$$\begin{aligned}\Delta\theta &= \frac{p(p + \omega_1)}{\omega_0\omega_1}\,\theta = \frac{p^2\theta + p\omega_1\theta}{\omega_0\omega_1} \\ &= \frac{p\theta}{\omega_0} + \frac{p^2\theta}{\omega_0\omega_1} = \frac{1}{\omega_0}\,\theta' + \frac{1}{\omega_0\omega_1}\,\theta''.\end{aligned} \tag{58}$$

Equation (58) states that the error $\Delta\theta$ may be regarded as made up of two parts, a velocity error $(1/\omega_0)\theta'$, which is proportional to the velocity θ' (of input or output), and an acceleration error $(1/\omega_0\omega_1)\theta''$, which is proportional to the acceleration θ'' (of input or output). The proportionality factor for the velocity error is $1/\omega_0$ and for the acceleration error is $1/\omega_0\omega_1$. Hence, if θ' and θ'' are regarded as determinable from the input function, then these two components of the error could be computed if the values of ω_0 and ω_1 or $\omega_0\omega_1$ were known. Turning now to the decibel gain curve (Fig. 10·3), Ferrell states that the values of the "intercept points" ω_0, ω_1 and $\sqrt{\omega_0\omega_1}$ can be determined as follows: The intersection of the straight line representing the low-frequency part of the curve with the zero db axis is ω_0; the intersection of the second segment, the "higher-frequency line" is $\sqrt{\omega_0\omega_1}$; and the value of ω at which the two lines intersect, the "corner-frequency", is ω_1. Thus if an asymptotic gain plot is available, the proportionality factors in Eq. (58) can be computed.

The smaller these proportionality constants the smaller will be the total error. Hence, large values of ω_0 and $\omega_1\omega_0$ will correspond to small values of the error. *Consequently, the further toward the high-frequency end of the curve that these intercept points occur the smaller will be the servo error.* This important correlation is the one proposed by Ferrell for use as an index of error of the system.

It is of interest now to determine how the intercept points ω_0, ω_1 and $\sqrt{\omega_0\omega_1}$ may be defined in terms of the relational parameters k_v, T, and ω_n introduced in Sec. 9·8 in our discussion of the transient response of a servo system. The two sets of parameters may be related by considering how the members of each set are defined in terms of physical or dimensional parameters. The dimensional parameters $S_m\mu_m$, R, and J used by Ferrell are equivalent to k_o, f, and J, respectively, the symbols used in Sec. 9·8. Ferrell's intercept parameters may therefore be translated into the previously introduced relational parameters as follows:

$$\begin{aligned} \omega_0 &= \frac{S_m\mu_m}{R} = \frac{k_0}{f} = k_v; \\ \omega_1 &= \frac{R}{J} = \frac{f}{J} = \frac{1}{T}; \\ \sqrt{\omega_0\omega_1} &= \sqrt{\frac{S_m\mu_m}{J}} = \sqrt{\frac{k_0}{J}} = \omega_n. \end{aligned} \tag{59}$$

Ferrell's parameters are shown in the first two columns of Eq. (59), and ours in the last two.[1] To complete the set of relations between the two sets of parameters, it is of interest to compute the value of the damping ratio ζ in terms of the intercept parameters.

$$\zeta = \frac{f}{2\sqrt{k_0J}} = \frac{1}{2}\frac{\omega_1}{\sqrt{\omega_0\omega_1}} \text{ or } \frac{1}{2}\sqrt{\frac{\omega_1}{\omega_0}}.$$

A performance property of considerable importance in many instrument servo systems is the steady-state velocity error. In Sec. 9·9, a method was described for computing it by means of the final value theorem, a procedure that may be considered a short-cut variant of the transient approach. Let us now consider how it might be determined by the application of Ferrell's intercept method. Consider first Ferrell's formulation, in Eq. (58), of the error equation for a second-order system. If, as is necessary in computing the steady-state velocity error, the input function is assumed to have a constant slope, say N, then at values of $t > 0$, $d^2\theta/dt^2$ equals zero, and Eq. (58) becomes

$$\Delta\theta = \frac{1}{\omega_0}N = \frac{N}{k_v}.$$

Thus the steady-state velocity error can be found simply by determining the value of k_v on the decibel-log ω diagram and dividing it into the slope N of the input function. Even where the decibel-log ω gain curve has been plotted from empirical data rather than a known feedback transfer function, it is possible to determine graphically the asymptote to the low-frequency end of the curve and thus determine its intersection with the 0-db axis, which will equal ω_0 or k_v.

[1] In Fig. 10·3 it may be noted that the intercept points are labeled in terms of both types of parameter, the designation in terms of Ferrell's symbols being given above the point and in terms of the k_v, T, ω_n set of parameters below the point.

The simplicity of the method suggests that it may prove an extremely useful procedure. It is therefore natural to inquire whether or not it can be used in the case of *higher-order transfer functions* as well as for the second-order system discussed by Ferrell.

Let the feedback transfer function of the system be given by Eq. (60) written in terms of the time-constant set of relational parameters. The order of the trans er function will equal the number of factors containing ω in the denominator minus the number in numerator. Only one factor is shown in the numerator in order to keep the expression as simple as possible, but this does not alter the logic involved.

$$\frac{\theta_o(j\omega)}{E(j\omega)} = \frac{k_v(j\omega T_a + 1)}{j\omega(j\omega T_1 + 1)(j\omega T_2 + 1) \cdots (j\omega T_n + 1)}, \qquad T_1 > T_2 \cdots > T_n. \tag{60}$$

$$\left|\frac{\theta_o(j\omega)}{E(j\omega)}\right| = k_v \frac{1}{\omega} \frac{1}{|j\omega T_1 + 1|} |j\omega T_a + 1| \frac{1}{|j\omega T_2 + 1|} \cdots \frac{1}{|j\omega T_n + 1|}$$

on the basis of the same line of reasoning followed in Sec. 10·3. At the low-frequency end of the decibel curve, that is, for

$$\omega \ll 1/T_1 < 1/T_2 \cdots < 1/T_n \text{ and } \omega \ll 1/T_a,$$

$$\left|\frac{\theta_o(j\omega)}{E(j\omega)}\right| \approx k_v \frac{1}{\omega}, \tag{61}$$

since all terms of the form $|j\omega T_k + 1|$ will approximately equal 1. Therefore

$$\left|\frac{\theta_o(j\omega)}{E(j\omega)}\right|_{\text{db}} \approx 20 \log k_v - 20 \log \omega. \tag{62}$$

Thus the first segment of the asymptotic gain curve is the same regardless of the number of factors of the form $(j\omega T_n + 1)$ in either numerator or denominator, since these all reduce to 1. Hence its intercept with the 0-db axis will be independent of the number of such factors in numerator or denominator. The value of this intercept is easily shown to equal k_v, by setting the left-hand side of Eq. (62) equal to 0.

$$\begin{aligned} 0 &= 20 \log k_v - 20 \log \omega \\ 20 \log \omega &= 20 \log k_v \\ \omega &= k_v. \end{aligned}$$

It is necessary now to show only that $1/k_v$ is always the velocity error constant of the system, i.e., that k_v is the constant which, when divided into the slope of the velocity input function, gives the steady-state velocity error. This may be done by computing the steady-state velocity error E_v in the usual way, from $\theta_o(s)/E(s)$. This transfer function is given by Eq. (63) with s substituted for $j\omega$.

$$KG(s) = \frac{\theta_o(s)}{E(s)} = \frac{k_v(T_a s + 1)}{s(T_1 s + 1)(T_2 s + 1) \cdots (T_n s + 1)}. \tag{63}$$

Applying Eq. (12) (Sec. 9·5) and substituting N/s^2 for $\theta_i(s)$,

$$E(s) = \frac{1}{1 + KG(s)}\,\theta_i(s)$$

$$= \frac{1}{1 + \dfrac{k_v(T_a s + 1)}{s(T_1 s + 1)(T_2 s + 1) \cdots (T_n s + 1)}}\,\frac{N}{s^2},$$

$$E(s) = \frac{s(T_1 s + 1)(T_2 s + 1) \cdots (T_n s + 1)N}{[s(T_1 s + 1)(T_2 s + 1) \cdots (T_n s + 1) + k_v(T_a s + 1)]s^2}.$$

Applying the Laplace transformation final value theorem (assuming that the specification is met concerning absence of poles on the $j\omega$ axis or in the right-half plane),

$$E(t)_v = \lim_{s \to 0} \frac{ss(T_1 s + 1) \cdots (T_n s + 1)N}{[s(T_1 s + 1) \cdots (T_n s + 1) + k_v(T_a s + 1)]s^2}$$

$$= \frac{N}{k_v},$$

since all terms of the form $(T_k s + 1)$ approach one.

Thus we see that for the type of system represented above k_v is the velocity error constant regardless of the number of factors of the form $(T_k s + 1)$ contained in numerator or denominator.[1] By similar reasoning if the first factor in the denominator of Eq. (60) is $(j\omega)^2$ rather than $(j\omega)$, then $k_v = \omega^2$ and $E(t) = (0)N/k_v$. In this case, the velocity error constant is zero regardless of the value of k_v.

Decibel Phase-margin Diagram.—The phase-margin diagram is notable for the simple form that Nyquist's criterion assumes when it is reformulated for use with this type of diagram. Figure 10·18 shows a family of curves representing a proportional servo system plotted for different values of the damping ratio ζ. The system is that defined by the transfer function of Eq. (21) (Sec. 9·7). In terms of Nyquist's criterion, any given curve indicates an unstable system if, as ω increases, the curve crosses the zero phase-margin axis (Y-axis) before it crosses the 0-db axis (x-axis). Any given curve is regarded as proceeding from above downward, i.e., from low to high values of ω. If the curve crosses the 0-db axis before reaching the zero-phase-margin axis, its phase

[1] This statement should not be taken to mean that the value of k_v will be the same regardless of the number and kind of energy-storage components or phase-advance components in the physical system. It means only that once k_v has been correctly determined for the over-all system in the way described in Sec. 10·3, then it can be regarded as the velocity error constant of the system.

margin at 0 db can immediately be read from the graph. Or if it is an unstable system, its phase-margin deficiency is equally directly perceptible.

In the provision of indices of accuracy, the phase-margin diagram is of no particular value. For this purpose the decibel-log ω diagrams described above should be used.

10·5. Operations on Frequency Diagrams.—In the present section are collected the chief operations that may be performed on the various types of frequency diagram. These operations are utilized in the design of corrective devices for compensating for system deficiencies. Some of them also find a use as graphical substitutes for computation.

Operations on Transfer Loci: Graphical Computation of the Output and Error Transfer Functions. The operations represented in Eqs. (9·12), and (9·14), giving output and error transfer functions in terms of $KG(j\omega)$, may be carried out graphically.[1] The only requirement is a plot of the transfer locus $KG(j\omega)$. Consider first the procedure used to find the output transfer function $\theta_o(j\omega)/\theta_i(j\omega)$. Figure 10·1 shows a representative transfer locus, with a vector drawn to the point C, corresponding to the frequency ω_c. Now

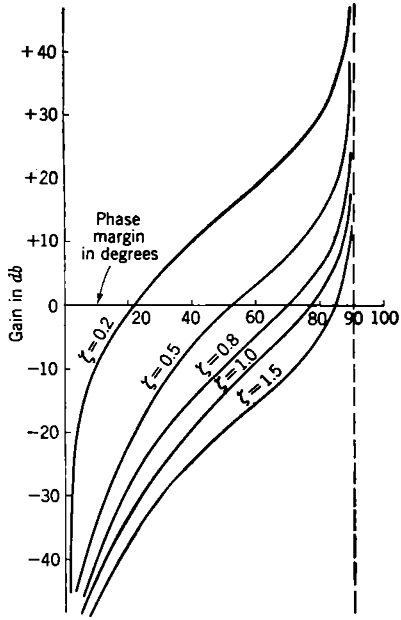


FIG. 10·18.—Phase-margin diagrams for a proportional servo system. Different curves differ in value of the damping ratio ζ. As ζ decreases, the phase margin decreases for a given gain.

$$\frac{\theta_o(j\omega)}{\theta_i(j\omega)} = \frac{KG(j\omega)}{1 + KG(j\omega)}.$$

The vector representing $KG(j\omega)$ is OC, since by definition, all points on the locus represent the function $KG(j\omega)$. If we add the vector 1 (represented by AO) to $KG(j\omega)$, we obtain

$$1 + KG(j\omega) = \text{AO} + \text{OC} = \text{AC}.$$

[1] See H. Harris, "The Analysis and Design of Servomechanisms," NDRC Report, 1942, pp. 51*ff.*; and A. C. Hall, *Analysis and Synthesis of Linear Servomechanisms*, Technology Press, Massachusetts Institute of Technology, 1943, pp. 30–33, for further details.

Therefore, $|KG|/|1 + KG| = |\mathrm{OC}|/|\mathrm{AC}|$; consequently, $|KG|/|1 + KG|$, which is equal to $\left|\dfrac{\theta_o(j\omega)}{\theta_i(j\omega)}\right|$ may be found by dividing the length of OC by the length of AC.

The phase of $\theta_o(j\omega)/\theta_i(j\omega)$ = phase angle of $KG(j\omega)$ minus phase angle of $(1 + KG) = \alpha - \beta = \delta$. The sign of δ will be negative. Thus the phase of $\theta_o(j\omega)/\theta_i(j\omega)$ is given in magnitude by the angle between the vectors representing $KG(j\omega)$ and $1 + KG(j\omega)$ and is negative in sign.

The same procedure, carried through for a *range* of values of ω, permits the frequency-response data for the amplitude and phase curves of $\theta_o(j\omega)/\theta_i(j\omega)$ to be obtained. These curves may be sketched in approximately simply by inspection of the transfer locus, or they can be determined more precisely by use of a protractor for measuring angles and dividers and ruler for measuring vector lengths.

A similar procedure can be used for graphical computation of the *error transfer function* $E(j\omega)/\theta_i(j\omega)$, as shown by Fig. 10·1. At any angular frequency ω_c,

$$\frac{E(j\omega)}{\theta_i(j\omega)} = \frac{1}{1 + KG(j\omega)},$$

$$\left|\frac{E(j\omega)}{\theta_i(j\omega)}\right| = \frac{1}{|1 + KG|} = \frac{1}{\mathrm{AC}}.$$

The phase of $E(j\omega)/\theta_i(j\omega) = 0 - \beta = -\beta$. This procedure carried through for the range of value of ω gives the necessary data for plotting the function $E(j\omega)/\theta_i(j\omega)$.

A second method for graphical determination of the output transfer function $\theta_o(j\omega)/\theta_i(j\omega)$ from the transfer locus depends on the plotting of transfer locus curves for which $\theta_o(j\omega)/\theta_i(j\omega)$ is a constant. On a transfer locus plot, these curves are circles whose radius and position are a function of the constant, to be designated as R. Figure 10·19, taken from Harris[1] shows the family of circles corresponding to different values of R. A similar family of curves exists for the phase angle ϕ.[2] If the family of curves of constant R are superimposed on the transfer locus A of a particular system, then the amplitude function of $\theta_o(j\omega)/\theta_i(j\omega)$ for that system can be determined from the points where locus A intersects the curves of constant R. For any given intersection, the value of ω is given by the ω of that point on the locus A. The corresponding value of the amplitude ratio $\theta_o(j\omega)/\theta_i(j\omega)$ is given by the R of that particular circle. The same procedure is used in finding the coordinates of the intersections with all other circles. One thus assembles a series of pairs

[1] Harris, *op. cit.*, p. 51*a*. Harris uses the symbol (GH) in place of the symbol used here of KG for the feedback transfer function.

[2] *Ibid.*, p. 55*a*.

of values representing $\theta_o(j\omega)/\theta_i(j\omega)$ as a function of ω. The same procedure carried out with the curves of constant phase superimposed on the locus A permits the phase-response curve for $\theta_o(j\omega)/\theta_i(j\omega)$ to be determined.[1]

At this point, it is convenient to indicate a way in which the circles of constant M may be used, in conjunction with the transfer locus, to indicate the frequency at which a peak will occur in the $\theta_o(j\omega)/\theta_i(j\omega)$

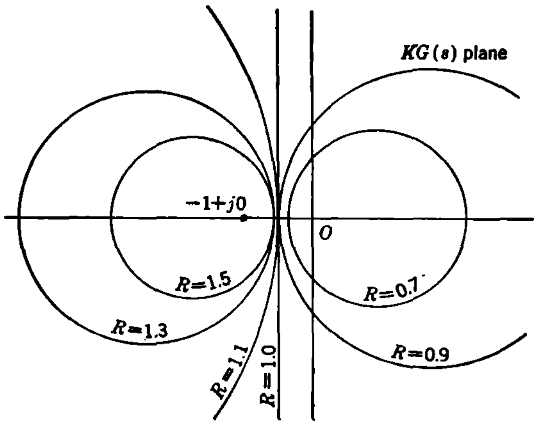


FIG. 10·19.—Transfer locus corresponding to constant $|\theta_0(j\omega)/\theta_i(j\omega)|$ ratios. (*Based on Harris, "The Analysis and Design of Servomechanisms," NDRC Report*, 1942, *Fig.* 25.)

amplitude curve and the height of this peak.[2] The transfer locus $KG(j\omega)$ is first plotted, and the family of circles of constant M superimposed upon it. Then the point of tangency of the circle that is tangent to the locus will indicate the frequency of either a maximum or minimum of the amplitude curve. The magnitude of the amplitude relation at this point will be given by the M characteristic of that circle or can be found by dividing the length of the $KG(j\omega)$ vector by that of the $[1 + KG(j\omega)]$ vector in the manner described some paragraphs earlier (i.e., in Fig. 10·1 OC is divided by AC). Whether this point gives a maximum or a

[1] The derivation of the formulas used in plotting the curves of constant amplitude may be found in Hall, *op. cit.*, pp. 50*ff.* Expositions of this method are given by both Harris, *op. cit.*, pp. 51*ff.*, and Hall, *op. cit.*, pp. 50–54. Hall uses the symbol M for the constant-amplitude ratio in place of R. His formulas for the points used in plotting the circles of constant amplitude are $c = -M^2/(M^2 - 1)$, where c is the number specifying location of the center of circle in the complex plane, and $r = M/(M^2 - 1)$, where r is the radius of the circle. The formula for c shows that the center of the circle will lie on the negative real axis when $M > 1$ and on the positive real axis when $M < 1$.

[2] A description of the method is given by Hall, *op. cit.*, pp. 49–52.

minimum can be easily determined by finding the amplitude ratio for a few points adjacent to the point of tangency, by either of the graphical procedures already described [i.e., by observing the M's associated with intersections of locus and circles at adjacent points or by graphical division of KG and $(1 + KG)$ vectors at these points]. The frequency and height of the peak is thus easily determined.

By inverting the method it can be used to determine the value of K, in $KG(j\omega)$, that will correspond to some specified value of the peak in the $\theta_o(j\omega)/\theta_i(j\omega)$ amplitude curve. One plots the circle with an M value corresponding to the height of the peak required (e.g., $1\frac{1}{3}$) and then adjusts the value of the gain factor K until the locus corresponding to this K is tangent to the circle.[1]

Operations on Transfer Loci: Scale Changes.—An operation of considerable importance in many problems is the determination of optimal values of the gain K in the transfer function $KG(j\omega)$. Since K is a constant and independent of frequency, the effect of changes in it may be shown either by plotting $KG(j\omega)$ for different values of K or by plotting the locus of $G(j\omega)$ and regarding changes in the gain factor K as corresponding to changes in the magnitude of the scale units of the real and imaginary axes. Each of these points of view is used in different procedures for determining the optimal value of K in servo-system adjustment.

Multiplication of Loci.—Another operation that will be found to be important in procedures for adjustment of servomechanisms is that of locus multiplication. It corresponds to the analytic operation of multiplying two transfer functions and the physical operation of connecting two networks in cascade. Figure 10·20 shows two loci, A and B, each of which can be assumed to represent the transfer function of two units connected in cascade. How may the locus of the over-all system $A \cdot B$ be obtained from their individual loci? The appropriate procedure follows directly from the fact that each locus represents a set of vectors and that any vector (corresponding to a complex number) stands for the transfer function of a given physical component at a particular frequency. Multiplication of loci is therefore equivalent to multiplication, at *each* of a number of angular frequencies, of the vectors (or complex

[1] Hall also gives two other methods of finding the optimum K, i.e., the value of K corresponding to a specified peak in the $\theta_o(j\omega)/\theta_i(j\omega)$ amplitude curve. One method consists in drawing various loci corresponding to different K values and finding for each locus the magnitude of the amplitude peak by the graphical division method. The K corresponding to the required peak is finally determined by interpolation. The other method, which he considers the simplest of the three, is based on the plotting of the $G(j\omega)$ locus instead of the $KG(j\omega)$ locus. The details are given by Hall, *op. cit.*, pp. 52*ff*. It seems to the present writer that the simplest method now available for adjusting K is that given in relation to the decibel-log ω method of plotting later on in this section.

numbers) corresponding to that frequency on the two curves. Such multiplication is carried out by arithmetic multiplication of the lengths of the vectors to give the absolute value of the product and addition of the angles of the two vectors. The nature of the resultant locus may be sketched in approximately by inspection or determined more exactly through the aid of ruler and dividers. Locus C in the figure represents the product of loci A and B.

Division can be carried through by the inverse process. Division of locus A by locus B will therefore be carried out, at any given frequency, by dividing the length of vector A by that of B to give the resultant length of the quotient and by subtracting the angle of vector B from that of A to give the phase angle of the quotient.

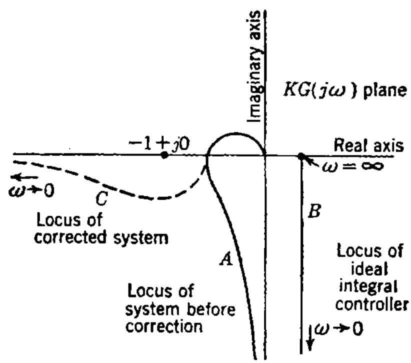


FIG. 10·20.—Correction of system with velocity log error (locus A) by means of integral controller (locus B). The locus of the corrected system C equals A times B.

Operations on Decibel-log ω and Phase-margin Diagrams.—The chief operations that may be performed on transfer loci, described above, find a parallel in operations that may be performed upon decibel-log ω diagrams. The starting point in both cases is, of course, the nature of the operation carried out on the transfer function.

Changes in the constant-gain factor K in the feedback transfer function $KG(j\omega)$ are represented by changes in the vertical level of the decibel-log ω gain curve relative to the 0-db axis. A given curve moves up as the gain increases and down as it decreases, the amount of change corresponding to the change in gain in decibels.

Since the gain factor K is independent of frequency, changes in it must alter the gain or amplitude function equally at all values of ω. Hence there can be no change in the shape of the gain curve, but only in its level.

A convenient way of determining how changes in the constant-gain factor K will influence performance properties of the system is provided by plotting the gain curves and phase curves on separate pieces of paper. Thus if the gain curve corresponding to a constant-gain factor of 1 (or 0 db) is plotted on a transparent piece of paper, the phase curve and coordinate scales on a second graph, and the first laid over the second, then the height of the gain curve may be readily shifted to correspond to different values of K. As the level of the gain curve changes with K, its relation to the phase curve will change to provide measures of system stability, as described in the previous section. The phase curve will not change with K, however, since it depends only on the shape of the gain curve and not on its level. The shifts in level of the gain curve may also be used to determine the effects on system accuracy, as represented by the intercepts of the gain curve with the 0-db axis (as discussed in Sec. 10·4). This method for determining the optimum gain K compatible with system stability will probably be found easier than that involving operations on the transfer locus intended to provide a specified peak in the $\theta_o(j\omega)/\theta_i(j\omega)$ amplitude curve.

Multiplication of two transfer functions can be carried out on the decibel-log ω diagrams by *addition* of the *gain curves* corresponding to the two functions to give the gain curve of the product and *addition* of the *phase curves* to give the phase curve of the product. The correctness of this procedure can be demonstrated by the familiar procedure of taking logarithms of the product. Thus, let

$$K_TG_T(j\omega) = K_aG_a(j\omega)K_bG_b(j\omega)$$

where $K_aG_a(j\omega)$ = transfer function of unit a,
$K_bG_b(j\omega)$ = transfer function of unit b,
$K_TG_T(j\omega)$ = transfer function of the units a and b connected in cascade.

The K symbols represent constants, and $G(j\omega)$ symbols represent the frequency-dependent parts of the transfer functions. Then

$$20 \log |K_TG_T(j\omega)| = 20 \log |K_aG_a(j\omega)| + 20 \log |K_bG_b(j\omega)|.$$

This equation states that if the functions corresponding to the first and second terms on the right-hand side are plotted separately, their sum will be the function on the left. But the decibel gain curves constitute the graphs of these functions; hence the gain curve of the product can be found by adding the gain curves of the components.

The rule for adding *phase* curves can be derived by writing the various transfer functions in exponential form and taking logarithms as before. Thus, if the subscripts have the same meaning as above and each transfer function is written in the form $Re^{j\phi}$, where R represents the amplitude ratio, ϕ the phase angle, and R and ϕ are each regarded as functions of $(j\omega)$, then

$$R_Te^{j\phi_T} = R_ae^{j\phi_a}R_be^{j\phi_b}.$$

Taking the natural logarithm of both sides,

$$\begin{aligned}\ln R_T + j\phi_T &= \ln R_a + j\phi_a + \ln R_b + j\phi_b \\ &= \ln R_a + \ln R_b + j(\phi_a + \phi_b).\end{aligned}$$

Equating real and imaginary parts of both sides,

$$\begin{aligned}\ln R_T &= \ln R_a + \ln R_b, \\ j\phi_T &= j(\phi_a + \phi_b);\end{aligned}$$

that is, $\phi_T(j\omega) = \phi_a(j\omega) + \phi_b(j\omega)$. This final equation is equivalent to the statement that the separate phase curves of components a and b, when added together, give the phase curve of the two units in cascade.

The operation for *division* of two transfer functions can be inferred immediately from the procedure for multiplication. Since the log of the quotient of two numbers is the log of the dividend minus the log of the divisor, an equivalent rule will hold for the quotient of two transfer functions. Hence, to obtain the *amplitude* function of the quotient, one *subtracts* the gain curve of the divisor transfer function from that of the dividend; to obtain the *phase curve* of the quotient, the phase curve of divisor function is subtracted from that of dividend function. That is to say, the rule for multiplication is used, but with the amplitude and phase curves of divisor transfer function given a negative sign.

If we turn now to the *phase-margin diagram,* we find that the procedure of representing changes in *gain* of the transfer function by changes in height of the curve still holds. For the phase curve associated with a given gain curve depends only on the shape of the gain curve and not on its level or distance from the 0-db axis.[1] Hence, corresponding to a gain curve of a given shape, there will be a unique phase curve and therefore a unique phase-margin diagram. Changing the gain factor of the transfer function will alter merely the level of the gain curve and not its shape. Hence there will be no change in the phase curve or the shape of phase-margin diagram. Changes in gain will thus be represented only by a constant change in the decibel coordinate of the phase-margin diagram for all values of ω. That is to say, there will be a change only in the height of the curve as a whole. Graphical multiplication and division of transfer functions by means of phase-margin diagrams may be carried out by procedures analogous to those used for multiplication and division of transfer loci. It is desirable, however, to use the phase-*angle* reference axis rather than the phase-*margin* reference axis. Then, if the phase-margin curves representing the different transfer functions are plotted, the phase-margin curve representing their product can be obtained by adding the ordinates (gain in decibels) corresponding to a particular value of ω on the two curves and, similarly, by adding the abscissas representing *phase angle.* This procedure carried out for the necessary range of values of ω provides the data for the over-all phase-margin curve. The same information may be obtained more easily, however, from the decibel-log ω curves.

The operations required for *graphical computation* of the $\theta_o(j\omega)/\theta_i(j\omega)$

[1] This conclusion follows from the relations established by Bode between the gain and phase functions of minimum phase systems. See H. W. Bode, "Relations between Attenuation and Phase in Feedback Amplifier Design," *Bell System Tech. Jour.*, **19**, 421–454, July 1940.

function from the feedback transfer function can also be carried out with the phase-margin diagrams. The procedure is directly analogous to that involving families of circles corresponding to constant $|\theta_o(j\omega)/\theta_i(j\omega)|$ ratios and constant phase relation between θ_o and θ_i. The same data [amplitude ratio and phase of $\theta_o(j\omega)/E(j\omega)$] represented in plots of these circles as transfer loci on a complex plane may be used in plots upon a phase-margin diagram. The resultant curves are no longer circles, but they may serve the same function as before. Each curve in one set corresponds to a constant ratio $|\theta_o(j\omega)|/|\theta_i(j\omega)|$. Each curve in a second set corresponds to a constant phase angle ϕ equal to arc $\theta_o(j\omega)$ $-$ arc $\theta_i(j\omega)$. The intersections of these curves with the phase-margin curve representing any given system provide the data for determining the amplitude and phase response curves for $\theta_o(j\omega)/\theta_i(j\omega)$.[1]

The other graphical method described in the early part of this section for computing $\theta_o(j\omega)/\theta_i(j\omega)$ and $E(j\omega)/\theta_i(j\omega)$ functions from $KG(j\omega)$ cannot be applied to decibel frequency diagrams, since the method involves addition of vectors and not solely multiplication or division. Logarithmic plots do not provide any equivalent for addition other than previous addition of the magnitudes themselves, since on a logarithmic diagram addition of curves is used to represent multiplication of the original functions.

[1] Charts for carrying out this method may be found in Vol. 25.

CHAPTER 11

SERVO THEORY: EVALUATION AND CORRECTION OF SYSTEM PERFORMANCE AND SPECIAL PROBLEMS

By G. L. Kreezer and I. A. Greenwood, Jr.[1]

EVALUATION OF SYSTEM PERFORMANCE

Once the performance properties of a given system have been determined by such methods as are reviewed in the preceding chapters, there arises the question of whether or not the performance meets specifications. The present section is concerned with ways available for thus evaluating a system. The notion of evaluation implies comparison of a test object with standards. Since no conventional set of standards of servo-system performance seems to have been set up, the standards adopted depending rather on the nature of the specific problem, the discussion may be limited to a brief survey of the kinds of performance property that are important and to an itemization of different ways of making the comparison of performance with a standard. These procedures are fairly obvious and may be reviewed briefly. Three methods may be mentioned: evaluation of the system on the basis of response curves, evaluation on the basis of a set of specifications of required performance properties, and evaluation by way of a single figure of merit for the entire system.

11·1. Response Curves.—On the graphical plot of the error or output time functions, lines or curves may be drawn indicating the allowable range within which the response curves may lie, thus making readily perceptible in just what region of the curve the standards of performance are not met. It may be specified, for example, that the transient or steady-state error for a given type of input should not exceed 1 per cent of the input after a given interval of time. Horizontal lines drawn at appropriate distances above and below the time axis of the error curve will show at a glance if this requirement is met. Similar procedures may be used to represent graphically other specifications of the required response.

11·2. Specifications for a Set of Performance Properties.—The required performance, instead of being represented on a graph, may be

[1] Sec 11·12 is by I. A. Greenwood, Jr.; the rest of Chap. 11 is by G. L. Kreezer.

formulated numerically in terms of the important properties of performance, classifiable under the heads of stability and accuracy.

Stability.—A servo system is said to be unstable if it shows oscillations that do not finally damp out.[1] This condition might be represented either by constant amplitude oscillations or by oscillations of steadily increasing amplitude. In any physical system the latter condition cannot continue beyond a certain point. After the amplitude passes a certain magnitude, regions of nonlinearity or approach to cutoff points will be reached in one or more of the system elements and the amplitude of the oscillations will fail to increase further. This transition to a nonlinear region will not, however, be indicated by the mathematical solution of a system assumed to be linear. Even though a system is found to be stable, in the sense specified above, its performance may be unacceptable if there are positively damped oscillations present that die out too slowly.[2] The specifications with respect to stability may be given in terms of the allowable number of oscillations or "overshoots" before the oscillations fall below a given amplitude; in terms of the magnitude of the real part of the complex root, which determines the rate of damping; in terms of the logarithmic decrement shown by successive cycles; or in terms of the damping ratio ζ for certain types of system. From the point of view of the frequency approach, stability requirements may be specified in terms of phase and gain margins.

Accuracy.—The allowable *transient error* may be specified in terms of the interval of time within which the error must fall to a given absolute magnitude or to a certain percentage of the input signal or in terms of the required time constant. The latter specification is equivalent, in the case of a simple exponential error curve, to the requirement that the error fall to 36.8 per cent of its initial value in the time specified for the time constant.

Under the head of *steady-state* errors, those of most interest are the displacement or static error and the velocity lag error. The standards in these respects may be given simply as the maximum allowable magnitudes of these two quantities.

11·3. Unitary Figures of Merit.—Some attempts have been made to provide some unitary measure of the "goodness" of a servo system in order to indicate its over-all value without restriction to a particular

[1] In mechanics the designation of a system as unstable is limited to those which give rise to oscillations or phenomena of progressively increasing magnitude. Such phenomena, in a linear system, correspond to a characteristic equation with roots that lie in the right half plane. In physical servo systems, it is customary to designate as unstable systems exhibiting oscillations of constant amplitude, even though in a strict mathematical sense they might be regarded as stable.

[2] This question of the rate at which oscillations damp out might possibly be considered more properly a problem of the transient error rather than of stability.

type of input function. Thus Hazen[1] and Brown[2] have proposed ways of obtaining such over-all ratings of the system. More elaborate mathematical procedures have been discussed by Phillips[3] and by Hall[4] on the basis of the minimum integral squared or mean squared error. In terms of this criterion, the best system or adjustment of a system is considered to be that which makes the following function of the error curve a minimum:

$$I = \int_0^\infty [E(t)]^2\, dt. \tag{1}$$

CORRECTION OF SERVO-SYSTEM PERFORMANCE

On the basis of the principles covered to this point, one may determine the response of a given system to various types of inputs and determine how this response compares with standards or specifications. Suppose the computed performance is deficient in some respect. How can the system be adjusted or corrected so as to eliminate these deficiencies? The present section of the chapter will deal with this question of servo-system correction. The problem can often be conveniently fractionated. One may determine first the *types* of device that can be used to eliminate a particular type of deficiency and then go on to determine the quantitative adjustments necessary in a given device so that the system will meet specifications. We shall be concerned with both types of problem. The transient method of analysis is often better adapted to the first type of problem, helping to give one an insight into the appropriateness of a particular kind of corrective network; the frequency method of analysis, on the other hand, is usually more effective in the problem of specific design. We shall have occasion to make use of both methods; no attempt will be made to carry both through completely on all problems. Interest will center rather in illustrating the ways in which each method can contribute to the general question of system correction. The survey of system deficiencies will deal with the same performance properties itemized in Sec. 11·2. Corrective procedures for stabilization will be considered first, followed by procedures for improvement of system accuracy.

[1] H. L. Hazen, "Theory of Servo-mechanisms," *Jour. Franklin Inst.*, **218**, 3, 322*ff*., September 1934.

[2] G. S. Brown, *Transient Behavior and Design of Servomechanisms*, privately printed, Massachusetts Institute of Technology, 1943 and 1945, p. 14.

[3] R. S. Phillips, "Servo Mechanisms," RL Internal Report No. 81-6, May 11, 1943, pp. 1–32.

[4] A. C. Hall, *Analysis and Synthesis of Linear Servomechanisms*, Technology Press, Massachusetts Institute of Technology, 1943, pp. 19–27.

STABILIZATION PROCEDURES

Mathematical analysis or empirical test may show the servo system to be unstable or to possess an insufficient stability margin. How can its stability be increased? Three methods have been found particularly useful: (1) introduction of phase advance[1] or a derivative error controller; (2) feedback of the derivative of the output signal, a method commonly known as tachometer feedback; and (3) the use of oscillation dampers. We wish to determine, in terms of the methods developed in the sections on transient and frequency analysis, why these devices are effective and how the correct parameters for a given device may be determined.

11·4. Derivative Error Controller (Phase Advance). *Transient Analysis.*—The simplest type of continuous controller is one in which the transfer function is a constant, as in the proportional servo. This type of system has been analyzed in Sec. 9·8. If the performance requirements are not too stringent, this simple system may prove satisfactory. In certain applications, however, it shows defects that require the development of a more complex type of controller.

The defects arise from the following two circumstances:

1. In the proportional controller the parameter that is effective in preventing oscillation, if it is sufficiently large, is the viscous friction f. But this parameter also involves the dissipation of energy in the system by heat. Unlike inertia or elastance parameters in the output load, it does not merely involve temporary storage of energy which is subsequently returned to the system. Consequently, if this damping term must be large in order to stabilize the system, on account of the other special properties of the system (such as a large inertia J or a large gain factor k_0), then there will be a large power loss that will be of use only for stabilizing the system. This loss is not serious, however, for small instrument servos.
2. A second type of defect arises when one needs to utilize an f of appreciable magnitude for damping the system and yet needs to bring the velocity lag error under a given level. The magnitude of the velocity lag error is proportional to f as shown by Eq. (3). It is impossible, therefore, both to increase f for damping purposes and to keep the velocity lag error from increasing. These relations are summarized by Eqs. (2) and (3). Equation (2) shows that the damping ratio ζ increases directly with f, but Eq. (3) shows that the velocity lag error θ_v does so also.

[1] The terms "phase advance" and "phase lead" are used interchangably.

$$\zeta = \frac{f}{2\sqrt{k_0 J}};\tag{2}$$

$$\theta_v = \frac{f}{k_0} N,\tag{3}$$

where N is slope of input ramp function.

The so-called *derivative error controller* has been developed in order to provide for a damping term in the characteristic equation of a servo system that will not have the disadvantages just enumerated for the viscous friction parameter f. Equation (4), the error equation of a proportional system, shows the presence of a term fs in the characteristic function.

$$E(s) = \frac{Js^2 + fs}{Js^2 + fs + k_0} \theta_i(s).\tag{4}$$

The first two terms of the characteristic function are due to the denominator of $H(s)$; the last to $C(s)$. Without the fs term in the characteristic function, the system would be oscillatory.[1] In this instance this term is provided by the viscous friction present in the output load. If a term of like nature could be provided by means of the $C(s)$ member and the fs term then reduced to zero by making the viscous friction f equal to zero or negligible, we should have a means of stabilizing the system without the defects enumerated above. This result is achieved by constructing the controller member so it operates on the error in the manner indicated by Eq. (5).

$$T_c(t) = k_0 E(t) + k_1 \frac{dE(t)}{dt}.\tag{5}$$

The transfer function of the controller is readily obtained.[2]

$$\frac{T_c(s)}{E(s)} = C(s) = k_0 + k_1 s.\tag{6}$$

If in the denominator of Eq. (4) the new value of $C(s)$ is substituted for k_0, and the fs term originally provided by $H(s)$ is reduced to zero, we obtain

$$E(s) = \frac{Js^2}{Js^2 + k_1 s + k_0} \theta_i(s).\tag{7}$$

The characteristic equation, Eq. (8), now has a damping term $k_1 s$, which does not depend on viscous friction; consequently, the defects introduced by an fs term will be eliminated.

$$Js^2 + k_1 s + k_0 = 0.\tag{8}$$

[1] See, for example, the solution of the equation of a system in which f is equal to zero in Sec. 9·8.

[2] The method has been given in Sec. 9·3.

In addition to the introduction of a term corresponding to the first derivative of the error, controllers have been proposed that introduce, in addition, a term corresponding to the second derivative of the error. It may have either a positive or negative sign. If the sign is positive, this new term has the effect of increasing the inertia of the system; if it is negative, of decreasing the effective inertia and hence of contributing to system stability. For discussions of this type of controller from the transient point of view, the reports of Minorsky,[1] Brown,[2] and Harris[3] should be consulted.

The method described above may be used as a basis for formulating a *general procedure* for determining the form of the controller transfer function required to correct a given system. The steps required may be outlined as follows:

1. Write the equation of the error function $E(s)$, with the controller represented by $C(s)$.

   In the problem above, with f assumed to be zero and $\theta_i(t)$ a step function,

$$E(s) = \frac{Js^2}{Js^2 + C(s)}\,\frac{1}{s} = \frac{Js}{Js^2 + C(s)}. \tag{9}$$

2. Decide on the form of the error function or, more specifically, of the characteristic equation necessary to produce a time solution with the required properties.

   In the present problem, the characteristic equation must have the form given by Eq. (8), with a sufficiently large damping parameter k_1.
3. Write the transfer function $C(s)$ in a form that meets the requirements of Step 2.

   In this problem, it is necessary in Eq. (9) that $C(s) = k_0 + k_1 s$, as should be apparent from comparison of Eq. (8) with the characteristic equation of Eq. (9).
4. Design a physical controller that will have the transfer function specified for $C(s)$.

   In this problem, the required controller is provided approximately by the networks of Fig. 11·3.

   It will be found that this same general procedure can be applied to meet other special requirements.

[1] N. Minorsky, "Directional Stability of Automatically Steered Bodies," *Jour. Am. Soc. Naval Eng.*, **34**, No. 2, 280–309, May 1922.

[2] Brown, *op. cit.*, pp. 17–26.

[3] H. Harris, "The Analysis and Design of Servomechanisms," NDRC Report, 1942, pp. 19–22.

Frequency Analysis.—Let us consider now how the problem of system stabilization is attacked on the basis of the frequency approach, more particularly through the use of the transfer locus method of representation. Curves A, B, and C of Fig. 11·1 represent the loci of a proportional servo system with progressively decreasing values of gain K. The Nyquist criterion immediately shows that the system represented by A is unstable since locus A encloses the critical point $(-1,0)$. What corrective measures are possible? Inspection of the figure suggests two alternative procedures. The gain of the system may be reduced to give stable loci such as B or C. Or the whole locus or the part of it in the neighborhood of the critical point may be rotated in a counterclockwise direction so that the critical point will no longer be enclosed and the locus will show the proper phase and gain margins.

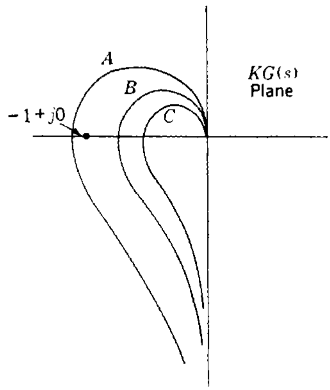


FIG. 11·1.—Effect of reduction in gain on transfer locus of unstable servo system.

System stabilization through a reduction in gain seems an attractive procedure, since it is relatively simple to carry out physically and, except for special types of system,[1] it should, if carried far enough, always be effective in producing stabilization. For as gain is reduced, the locus contracts until it finally no longer encloses the critical point.[2] Unfortunately, gain reduction cannot be relied on as a generally satisfactory procedure for stabilization, since it will also reduce the accuracy of the system by increasing the steady-state and transient errors. It may be used, therefore, only up to the point permitted by specifications of required accuracy. An early step required of the designer, therefore, is to compute the minimum gain required to meet accuracy specifications. Let us suppose that after this has been done, the gain requirement is such as to give rise to a locus of the form of A in Fig. 11·1 and that stabilization by further reduction in gain is therefore not feasible. We must therefore consider the second procedure proposed above.

[1] It cannot be depended on to produce stabilization in systems with loci of the type found in *conditionally stable systems* and in cases in which the physical nature of the components are such as to prevent gain reduction to be carried beyond a given point without introduction of new disturbing factors.

[2] Or alternatively, as suggested in Sec. 10·5, one may think of only the $G(j\omega)$ part of the feedback transfer function being plotted and the gain K represented by the scale value. As K decreases, the critical point will move to the left on the negative real axis and thus may be shifted to the left of the locus.

To rotate all or parts of a locus, such as curve A in Fig. 11·2, in a counterclockwise direction, it is apparent that the given locus must be multiplied by one with leading phase angles, such as given by Curves B, C, and D in Fig. 11·2. As indicated in Sec. 10·5, multiplication of two loci at any specific frequency means multiplication of the corresponding vectors, and this operation consists of addition of the phase angles and multiplication of the lengths of the vectors. Hence, to reduce the absolute value of any of the negative phase angles of Curve A, they must be added to positive phase angles, such as are provided by loci B, C, and D. Multiplication of the unstable locus A by phase-lead loci B, C, and D leads to loci B', C', and D', respectively. Nyquist's criterion shows immediately that the systems represented by the latter three loci will be stable.

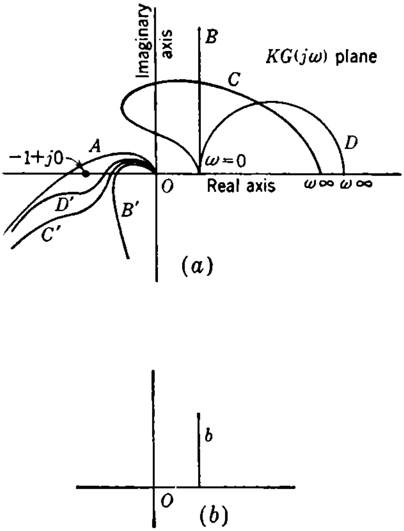


FIG. 11·2.—(*a*) Different types of phase-lead loci and their effect on the locus A of an unstable system; (*b*) transfer locus of a derivative controller.

At this point it is of interest to compare the results derived from the transient and from the frequency approach to the present problem of system stabilization. On the basis of the transient approach, it was concluded that an unstable proportional servo system could be stabilized by means of a derivative error controller, one with a transfer function of the form given by Eq. (6); on the basis of the frequency approach, it was concluded that the system could be stabilized by introducing in cascade devices possessing loci with a suitable range of positive phase angles as in loci B, C, and D of Fig. 11·2a. Are these two proposals two ways of saying the same thing, or do they point to different kinds of corrective device? An answer can be obtained by expressing both principles in the same terms. Let us plot the transfer function of the derivative error controller as a transfer locus. Substituting $j\omega$ for s in Eq. (6), we obtain

$$\frac{T(j\omega)}{E(j\omega)} = k_0\left(1 + \frac{k_1}{k_0}j\omega\right). \tag{10}$$

This transfer function can be plotted by inspection to give locus b of Fig. 11·2b. At $\omega = 0$, the transfer vector equals k_0; at $\omega = 1$, it equals

$k_0 + j(k_1/k_0)$; and so on. We see that locus b of the derivative error controller is identical in form with phase-lead locus B of Fig. 11·2a. The two approaches thus lead to consistent results. The frequency approach, however, suggests the suitability of a wider class of corrective devices (such as those corresponding to loci C and D of Fig. 11·2a) than has the transient approach. Locus b, orresponding to an ideal derivative network, can be considered a special case of the general class of loci showing a range of positive phase angles.

Can loci of this class be physically realized? It is commonly known that an *exact* synthesis of a derivative network is not possible. No passive network can be made to yield an output proportional to the mathematical derivative of the input, for all types of input. Consequently, a network

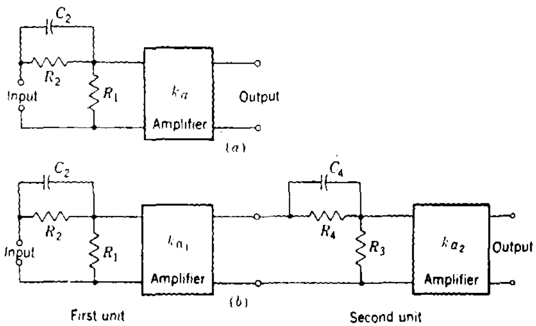


FIG. 11·3.—Circuit diagrams of (a) basic and (b) compound lead controllers.

cannot be built that will have a locus like that of B (Fig. 11·2a). Physical networks can be constructed, however, that will correspond to the other loci represented, such as C and D. Such loci can be used to bring about a counterclockwise rotation of the unstable locus (as A) in the region in which we are primarily interested, the neighborhood of the critical point. Networks capable of approximating a derivative network, in the sense that they provide for positive phase advance, can thus be realized and used as a basis for stabilization adjustments.

Circuit diagrams of physical networks that can be used for this purpose are given in Fig. 11·3. The circuit of Fig. 11·3a, designated by Hall as a basic lead controller, shows a passive network very commonly used for obtaining the approximate derivative of a signal, in cascade with an isolating amplifier. It corresponds to a locus of type D in Fig. 11·2a. Figure 11·3b shows a compound lead controller. Its transfer function is represented by a locus of type C in Fig. 11·2a. This type of controller is capable of providing a greater maximum phase advance than the basic lead controller and approximates a second derivative error controller. For a discussion of the factors to be taken into account in the selection of a particular type of lead controller and of methods for adjusting its parameters to produce the necessary amount of phase advance, the detailed report by Hall should be consulted.[1]

The stabilization principles discussed above in relation to the transfer locus may be reformulated for use with the *decibel-log frequency plots and decibel phase-margin diagrams*. It will be sufficient here, in view of the detailed treatment of the decibel method in the previous chapter, to show how a typical phase-advance controller is represented on a decibel plot and the effect of incorporating such a controller in a sample system.

Let us take the basic lead controller of Fig. 11·3 as our example. It is readily shown by the method described in Sec. 9·3 that the transfer function of the controller is given[2] by Eq. (11).

$$K_dG_d(s) = \frac{e_o(s)}{e_i(s)} = k_a \frac{R_1}{R_1 + \dfrac{R_2}{1 + R_2C_2s}} \tag{11}$$

$$= \frac{R_1k_a}{R_1 + R_2} \frac{1 + R_2C_2s}{1 + \dfrac{R_1R_2C_2s}{R_1 + R_2}}$$

$$= \frac{k_a}{\alpha_d} \frac{1 + \alpha_dT_ds}{1 + T_ds}, \tag{12}$$

where $T_d = R_1R_2C_2/(R_1 + R_2)$, the network time constant, and

$$\alpha_d = \frac{R_1 + R_2}{R_1},$$

the attenuation constant. Equation (12) gives the transfer function of the phase advance controller in a form suitable for plotting by the decibel approximation method previously described. Considering only the frequency dependent part, it is given as a frequency transfer function by

$$G_d(j\omega) = \frac{j\omega\alpha_dT_d + 1}{j\omega T_d + 1}. \tag{13}$$

The constant k_a/α_d will merely shift the gain curve upward by an amount equal to the constant, in decibels. A brief examination of Eq. (13) permits us to plot it by inspection. We note, first of all, that as

$$\omega \to 0, \ |G_d(j\omega)|_{\text{db}} \to 0.$$

The low-frequency asymptote will therefore coincide with the 0-db axis. The corner points will be at $\omega = 1/\alpha_dT_d$ and $\omega = 1/T_d$. Since α_d is greater than 1, $\alpha_dT_d > T_d$, thus indicating that the corner point corre-

[1] Hall, *op. cit.*, pp. 89–127.

[2] Hall, *op. cit.*, pp. 95*ff*. See also Harris, *op. cit.*, pp. 39*ff*.

sponding to $1/\alpha_d T_d$ occurs first, i.e., at the lower frequency. When the low-frequency asymptote reaches this corner point, with increasing frequency, it will slope up at 6 db per octave corresponding to the factor in the numerator of Eq. (13)] until it reaches $\omega = 1/T_d$, at which point it is added to a slope of -6 db per octave (corresponding to the denominator term), thus resulting in another horizontal line with a net slope of 0 db per octave. This gain curve is plotted as Curve *A* in the lower part of Fig. 11·4.

The upper gain curves of Fig. 11·4 show the effect of cascading the phase-lead controller with a proportional servo system. The broken line starting at $\omega = 1/\alpha_d T_d$ shows how the gain curve of the servo system (*B*), represented by the solid line, is altered by introduction of the lead network. The effect is to reduce the average rate at which the gain level falls with increasing frequency. This effect is important, since, as Bode[1] has emphasized, the magnitude of the phase lag is directly dependent on the rate of attenuation, increasing as the rate of attenuation increases. The phase-lead controller, by decreasing the average rate of attenuation in the frequency range in which gain level is greater than zero, decreases the net phase lag and so permits the gain crossover to occur before the phase lag has fallen to 180°. Consequently, a system that might otherwise be unstable becomes stable. In general, the introduction of a phase-lead network decreases the rate at which the phase curve approaches the phase crossover point by decreasing the average attenuation rate of the gain curve.

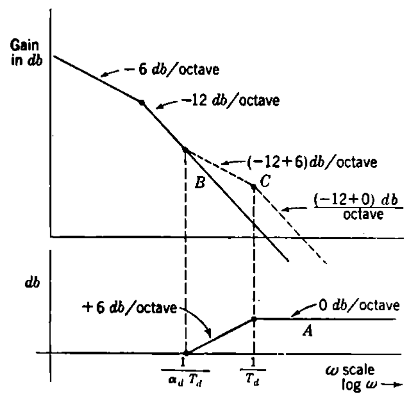


FIG. 11·4.—Effect of adding decibel gain curve (*A*) of derivative network to that of proportional servo system (*B*).

11·5. Derivative (Tachometer) Feedback.—In many systems it may be inconvenient to insert a network at a point lying between the error detector and the servomotor for the purpose of taking the derivative of the error, or the nature of the physical signal at accessible points may be such that a suitable differentiating device may not be constructed easily. In such cases, it is often possible to achieve the same effect by a device inserted at some other, more accessible point. Figure 11·5 is a block diagram of a servo system characterized by what is commonly called

[1] H. W. Bode, "Relations between Attenuation and Phase in Feedback Amplifier Design, *Bell System Tech. Jour.*, **19**, July 1940.

tachometer feedback. The tachometer feeds back to the input of the amplifier an electrical signal proportional to the rate of motion of the output displacement. If the sign of the feedback signal is positive, this signal will help to oppose the effect of any viscous friction present; if negative, it will supplement or substitute for viscous friction and thereby help to stabilize the system. How are these effects to be understood in relation to our previous analysis of the factors determining stability?

Consider first the error equation of a proportional servo system as given by Eq. (14),

$$E(s) = \frac{Js^2 + fs}{Js^2 + fs + k_0}\,\theta_i(s). \tag{14}$$

As pointed out in Sec. 9·8, if a term of the form of fs is absent from the denominator, the system will show oscillations of constant amplitude.

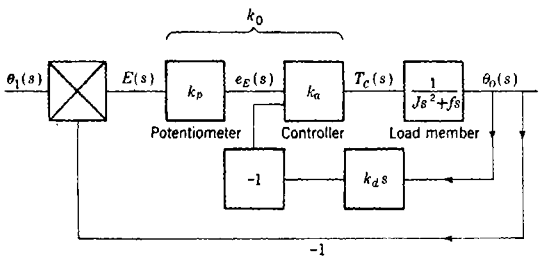


FIG. 11·5.—Proportional servo system with tachometer feedback.

The damping out of these oscillations can come about either through the presence of viscous friction f, which provides a term fs in the characteristic equation, or a network that takes the derivative of the error to provide a term k_1s of the same form. Similarly, a device that follows the output θ_o and feeds back a signal proportional to its derivative may also be used for the introduction of a term of this form.

This result is readily demonstrated analytically. The relevant transfer functions and the symbols for signals at various points are shown on the block diagram. The diagram represents the same proportional servo system considered previously, with the addition of a feedback link from θ_o to the amplifier input. The over-all gain factor k_0 is regarded as broken up into k_p, the conversion factor of a potentiometer for converting error angle into error voltage, and k_a, the conversion factor of the amplifier-motor combination. The dimensions of k_p will be volts per radian error and of k_a, torque in pound-feet per volt error. It will be noted that the feedback transfer function represents a device that takes the derivative of $\theta_o(t)$.

$$\left.\begin{aligned} e_F(t) &= k_d \frac{d\theta_o(t)}{dt}; \\ e_F(s) &= k_d s\theta_o(s). \end{aligned}\right\} \tag{15}$$

This operation is readily performed by a d-c electrical generator which, when mounted on the output shaft of the motor, will deliver a d-c voltage proportional to the shaft angular velocity. A (-1) conversion factor has been introduced after e_F to symbolize the fact that the feedback link is connected so that e_F will be subtracted from e_E rather than added to it. Some means are provided to control the size of the proportionality factor k_d.

Let us now find the feedback transfer function $\theta_o(s)/E(s)$. The necessary component equations are

$$\theta_o(s) = \frac{1}{Js^2 + fs} T_c(s), \tag{16}$$

$$T_c(s) = k_a[e_E(s) - e_F(s)], \tag{17}$$

$$e_E(s) = k_p E(s), \tag{18}$$

$$e_F(s) = k_d s\theta_o(s). \tag{19}$$

Substituting $e_E(s)$ and $e_F(s)$ as given by Eqs. (18) and (19) into Eq. (17), and substituting $T_c(s)$ as given by Eq. (17) into Eq. (16), we obtain

$$\theta_o(s) = \frac{1}{Js^2 + fs} [k_a k_p E(s) - k_a k_d s\theta_o(s)].$$

Solving for $\theta_o(s)/E(s)$,

$$\frac{\theta_o(s)}{E(s)} = \frac{k_a k_p}{Js^2 + (f + k_a k_d)s} = \frac{k_0}{Js^2 + (f + k_a k_d)s}. \tag{20}$$

Comparison of this equation with the feedback transfer function given by Eq. (9·21) shows that $(f + k_a k_d)$ here plays exactly the same role as f does in the system without tachometer feedback. This fact indicates, first of all, that the transient and frequency analyses given in Secs. 9·8 and 10·3 for a system with viscous friction will hold here too, with the new term substituted for f. It indicates, second, that we might permit the viscous friction f to equal zero and its role in stabilizing the system to be taken over by $k_a k_d$. Since k_a, the gain of amplifier-motor combination can be assumed to be given, k_d is the only parameter that would need to be controlled. It can be selected so that the product $k_a k_d$ will be equal numerically to whatever value of f is found necessary to suitably damp the system, since f and $k_a k_d$ have exactly the same dimensions (torque per radian-per-second).

The fact that the term $(f + k_a k_d)$ in Eq. (20) plays the same role as

f alone in the equation of a proportional servo (one with viscous friction in the load and no tachometer feedback) means that this term will also take the place of f in the equation for the velocity lag error. This error will consequently be given by

$$\frac{f + k_a k_d}{k_0} N \tag{21}$$

instead of by $(f/k_0)N$, as in Eq. (9·70). A simple method to prevent the tachometer feedback from contributing to the velocity error is to place a condenser in the feedback path between the tachometer and amplifier. The condenser may, in physical terms, be regarded perhaps as blocking the low-frequency signals that contribute to the steady-state velocity error and passing without interference the higher-frequency signals that help to stabilize the system. A more detailed discussion of the role of tachometer feedback loops in stabilizing a servo system may be found in Harris,[1] Williams,[2] and Jofeh.[3] Oscillation dampers, to be discussed in a later section, are also of interest here, since it can be shown that such dampers have exact analogues in different types of feedback network.

11·6. Oscillation Dampers.—A third method of stabilization useful in some systems makes it possible to change the relative amounts of inertia and viscous friction present in the load member through the medium of a special type of oscillation damper. This device, which may be of the tuned or untuned type, consists of a shell providing some additional inertia and an internal flywheel coupled to the shell by means of the viscous friction of the oil in which the internal flywheel is immersed. The shaft connected to the outer shell is coupled rigidly to the shaft of the load member. The damper as described above is of the untuned type. If, in addition, a spring is introduced to retard the movement of the internal flywheel relative to the case, the damper is said to be of the tuned variety. Although flywheel dampers have been widely used for stabilizing mechanical systems, the use of such dampers for purposes of servo-system stabilization is apparently due to the MIT Servomechanisms Laboratory. An analysis from the transient point of view has been given by Jofeh.[4] The frequency approach may also be used. Both points of view are considered below.

[1] H. Harris, "The Analysis and Design of Servomechanisms," NDRC Report, 1942, pp. 68–72.

[2] F. C. Williams, "Automatic Following Mirror Systems," TRE Report No. T-1505.

[3] L. Jofeh, "Shooting at Aircraft," Report No. 100, A. C. Cossor, Ltd.

[4] L. Jofeh, "Application of Oscillation Damper to Servo-mechanisms," RL No. Coss 3914; OSRD ref. No. WA-271324, Report No. MR-141, A. C. Cossor, Ltd., August 1944.

Untuned Damper: Transient Analysis.—Figure 11·6 shows the block diagram of a proportional servomechanism equipped with an untuned damper. The following four symbols have already been defined in the

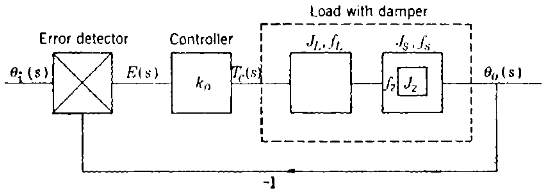


FIG. 11·6.—Block diagram of servo system with untuned viscous friction flywheel damper.

section on transient analysis, namely, $\theta_i(t)$, $\theta_o(t)$, $E(t)$, and k_0. The following additional symbols will also be used.

$\theta_d(t)$ = position of the damper flywheel relative to bearings of output shaft, i.e., measured from the same reference position used in measuring θ_o,

J_L = moment of inertia of the load without the damper attached,

f_L = viscous friction of the load without damper,

J_S = moment of inertia of the damper shell,

f_S = any additional viscous friction operating on initial load due to attachment of damper (this can ordinarily be assumed to equal zero),

J_2 = moment of inertia of internal flywheel of damper,

f_2 = viscous friction between outer shell of damper and inner flywheel,

$J_1 = J_L + J_S$. By means of this equation we can lump the inertia of original load and that added by damper shell.

$f_1 = f_L + f_S$. Similarly we can lump the viscous friction existing between output shaft and bearings of output shaft.

The following three equations describe the dynamics of the separate units of the system, namely, the error detector, controller, and load with damper, respectively.

$$E(t) = \theta_i(t) - \theta_o(t). \tag{22}$$

$$T_c(t) = k_0E(t). \tag{23}$$

$$T_c(t) = \left(J_1\frac{d^2\theta_o}{dt^2} + f_1\frac{d\theta_0}{dt}\right) + J_2\frac{d^2\theta_d}{dt^2}. \tag{24}$$

Equations (22) and (23) are merely a repetition of the relations with which we are familiar from our analysis of the proportional controller. Equation (24) represents the equilibrium of torques involved in the load member. The equation states that the applied torque $T_c(t)$ equals the opposing torques due to acceleration and velocity of different parts of the load. This opposing torque is made up of two parts: that provided by the part of the load which is rigidly coupled to the output shaft and that dependent on the internal structure of the damper. The first portion is represented by $J_1(d^2\theta_o/dt^2) + f_1(d\theta_o/dt)$ just as in the case of a load without an attached damper, but with the parameters of the damper shell lumped with those of the initial load. The second factor in the opposing torque may be represented by $\left(J_2 \frac{d^2\theta_d}{dt^2}\right)$, the opposing torque due to the acceleration of the internal flywheel. The viscous friction existing between the flywheel and the outer shell may be regarded as merely providing a coupling connection by means of which motion of the outer shell is transmitted to the inner flywheel and does not enter into the equation. The angular position of the inner mass will be less than that of the shell, lagging behind it by an amount $(\theta_o - \theta_d)$, representing the forward slip of the shell relative to the flywheel. If we consider the torques acting on the flywheel alone, we can write

$$f_2 \frac{d[\theta_0(t) - \theta_d(t)]}{dt} = J_2 \frac{d^2\theta_d(t)}{dt^2}. \tag{25}$$

This equation states that the torque applied to the inner flywheel is provided by the viscous friction existing between output shaft or shell and the flywheel and that it is opposed by the equal back torque associated with acceleration of the flywheel. This equation will be useful in permitting us to eliminate θ_d from our over-all servo equation.

On the basis of Eqs. (23) and (24) the transfer functions of the controller and output members may be obtained and combined to give the *feedback transfer function*. From it the *error transfer function* can be derived. The error function will provide the point of departure for the transient analysis considered below, while the feedback function may be used as the starting point of the frequency analysis. From Eq. (23) we obtain

$$\frac{T_c(s)}{E(s)} = k_0 \tag{26}$$

in torque units per radian error. This is the conversion factor of the controller, which here consists of error potentiometer, amplifier, and motor, in cascade. From Eq. (24), we obtain

$$T_c(s) = (J_1s^2 + f_1s)\theta_o(s) + J_2s^2\theta_d(s). \tag{27}$$

To eliminate $\theta_d(s)$ from Eq. (27) we make use of Eq. (25). Taking the Laplace transformation of this function and solving for $\theta_d(s)$, we obtain

$$\theta_d(s) = \frac{f_2}{J_2s + f_2}\,\theta_o(s) \tag{28}$$

Substituting for $\theta_d(s)$ in Eq. (27), we obtain

$$T_c(s) = (J_1 s^2 + f_1 s)\theta_o(s) + J_2 s^2 \frac{f_2}{J_2 s + f_2}\,\theta_o(s). \tag{29}$$

After some simplification we get

$$T_c(s) = \frac{J_1 J_2 s^3 + (f_1 J_2 + f_2 J_1 + f_2 J_2)s^2 + f_1 f_2 s}{J_2 s + f_2}\,\theta_o(s), \tag{30}$$

$$\frac{\theta_o(s)}{T_c(s)} = \frac{J_2 s + f_2}{s[J_1 J_2 s^2 + (f_1 J_2 + f_2 J_1 + f_2 J_2)s + f_1 f_2]}. \tag{31}$$

Multiplying Eqs. (26) and (31) there results

$$\frac{\theta_o(s)}{E(s)} = \frac{k_0(J_2 s + f_2)}{s[J_1 J_2 s^2 + (f_1 J_2 + f_2 J_1 + f_2 J_2)s + f_1 f_2]}. \tag{32}$$

This equation represents the feedback transfer function $KG(s)$. By applying Eq. (9·12) and a little algebra, we obtain the error transfer function $E(s)/\theta_i(s)$:

$$\frac{E(s)}{\theta_i(s)} = \frac{1}{1 + \dfrac{k_0(J_2 s + f_2)}{s[J_1 J_2 s^2 + (f_1 J_2 + f_2 J_1 + f_2 J_2)s + f_1 f_2]}} \tag{33}$$

$$= \frac{J_1 J_2 s^3 + (f_1 J_2 + f_2 J_1 + f_2 J_2)s^2 + f_1 f_2 s}{J_1 J_2 s^3 + (f_1 J_2 + f_2 J_1 + f_2 J_2)s^2 + (f_1 f_2 + k_0 J_2)s + f_2 k_0}. \tag{34}$$

Equations (32) and (34) can be written in terms of relational parameters by means of the following relations, the first two of which will be familiar from our treatment of the proportional servo system in Sec. 9·8:[1]

$$\left.\begin{aligned}
\frac{f_1}{J_1} &= 2\alpha_1, & f_1 &= J_1 2\alpha_1,\\
\frac{k_0}{J_1} &= \omega_{10}^2, & k_0 &= J_1\omega_{10}^2,\\
\frac{f_2}{J_2} &= 2\alpha_2, & f_2 &= J_2 2\alpha_2 = 2J_1 n\alpha_2,\\
\frac{J_2}{J_1} &= n, \quad \text{and} & J_2 &= J_1 n.
\end{aligned}\right\} \tag{35}$$

(35)[2]

[1] The relational parameters defined in Sec. 9·8 have the following relations to α and ω_{10}, as may be easily verified by inspection of the defining equations:

$$T = \frac{1}{2\alpha};$$
$$\omega_n = \omega_{10}.$$

[2] The symbols used in defining this set of relationships are those of Jofeh. Subsequently, however, in the representation of response relationships, the symbols introduced in the discussion of quadratic factors will be used.

If we make use of the relations given in the right-hand column, substituting them for f_1, f_2, k_0, and J_2 in Eq. (32), all the original physical parameters are readily eliminated, giving the transfer function in terms of relational parameters,

$$\frac{\theta_o(s)}{E(s)} = \frac{\omega_{10}^2 s + 2\alpha_2\omega_{10}^2}{s^3 + 2[\alpha_1 + \alpha_2(n+1)]s^2 + 4\alpha_1\alpha_2 s}. \tag{36}$$

The error transfer function, in terms of the same parameters, can be obtained by the usual procedure [i.e., application of Eq. (9·12)],

$$\frac{E(s)}{\theta_i(s)} = \frac{s^3 + 2[\alpha_1 + \alpha_2(n+1)]s^2 + 4\alpha_1\alpha_2 s}{s^3 + 2[\alpha_1 + \alpha_2(n+1)]s^2 + (4\alpha_1\alpha_2 + \omega_{10}^2)s + 2\alpha_2\omega_{10}^2}. \tag{37}$$

Equations (36) and (37) provide a basis for selection of appropriate constants for the damper in a given problem.

As Jofeh points out, two familiar types of problems may arise. One is to determine the response of the system, given the parameters of the system and the damper; the other is to determine the design parameters of the damper, given the parameters of the system without damper and specifications of required performance. The treatment of the first problem is straightforward and may follow the method described in the sections on the transient approach. It will be noted that the characteristic equation of Eq. (37) is a cubic. It may therefore be factored in the manner proposed by Liu and Evans [see Eq. (9·74)]. The characteristic equation of Eq. (37) can thus be represented as

$$(s + \xi\omega_{nq})(s^2 + 2\zeta_q\omega_{nq}s + \omega_{nq}^2) = 0. \tag{38}$$

If the parameters of this equation are known in terms of the given parameters of the system, the response characteristics can be found. The values of ζ_q and ω_{nq} and a set of nondimensional charts for the quadratic will give the response characteristics corresponding to the quadratic factor, and the real root $\xi\omega_{nq}$ will specify the characteristics of the exponential component. It becomes necessary then to determine the values of ζ_q, ω_{nq}, and $\xi\omega_{nq}$ in terms of the physical or relational parameters of the original system and damper. This may be accomplished by the familiar procedure of matching coefficients of two equations regarded as equivalent. Thus if the factors in Eq. (38) are multiplied out, we obtain

$$s^3 + \omega_{nq}(2\zeta_q + \xi)s^2 + \omega_{nq}^2(1 + 2\zeta_q\xi)s + \xi\omega_{nq}^3 = 0. \tag{39}$$

If the coefficients of this equation are set equal to the corresponding coefficients of the characteristic function in Eq. (37), the following set of equalities is obtained.

$$\left.\begin{aligned} \omega_{nq}(2\zeta_q + \xi) &= 2[\alpha_1 + (1+n)\alpha_2], \\ \omega_{nq}^2(1 + 2\zeta_q\xi) &= 4\alpha_1\alpha_2 + \omega_{10}^2, \\ \xi\omega_{nq}^3 &= 2\alpha_2\omega_{10}^2. \end{aligned}\right\} \tag{40}$$

All the parameters on the right-hand side of these three equations will be known from the data given concerning the system with damper; the left-hand side contains a total of three unknowns, ζ_q, ξ, and ω_{nq}. Since there are three equations, the values of the three unknowns can be found and hence the response characteristics of the system.

The second type of problem, that of determining design parameters for the damper, can be solved in a similar manner. In this case, the parameters of the original system and response specifications are given. Let us suppose that data on the original system are given in terms of physical parameters J_1 and f_1 but that gain k_0 is left unspecified. One relational parameter $2\alpha_1$ can be obtained from the ratio of f_1 to J_1. The specification of required response properties can be given in terms of parameters ζ_q, ω_{nq}, and $\xi\omega_{nq}$, since, as pointed out, specific response components correspond to these parameters. Referring now to the three equations in (40), we note that there will be four known parameters, the set just mentioned and three unknown parameters, namely, ω_{10}, α_2, and n. Consequently, the three simultaneous equations may be solved for these three unknowns. If we refer now to Eq. (35), it will be noted that the parameters thus solved for will determine the physical parameters of the damper and the constant-gain factor k_0, for the three relational parameters ω_{10}, α_2, and n are defined, by way of three equations, in terms of four physical parameters, k_0, J_1, J_2, and f_2. One of these, J_1, is one of the known properties of the original system. Hence the values of the other three physical parameters k_0, J_2, and f_2 can be computed. The problem of determining the design parameters for the damper is thus completed.

Tuned Damper: Transient Analysis.—An exactly parallel treatment can be carried out for the tuned damper. The detailed steps may be omitted and only the chief results given. The error-transfer function for the tuned damper is found to be

$$\frac{E(s)}{\theta_i(s)} = \frac{J_1J_2s^4 + (J_1f_2 + J_2f_1 + J_2f_2)s^3 + (f_1f_2 + S_2J_1)s^2 + S_2f_1s}{J_1J_2s^4+(J_1f_2+J_2f_1+J_2f_2)s^3+(f_1f_2+S_2J_1+k_0J_2)s^2+(S_2f_1+k_0f_2)s+k_0S_2}, \tag{41}$$

where the only new parameter is the elastance of the damper spring S_2. This equation is converted to one written in terms of relational parameters exactly as before. The only new relational parameter involved is ω_{20}^2, defined as S_2/J_2.

$$\frac{E(s)}{\theta_i(s)} = \frac{s^4 + 2[\alpha_1 + (n+1)\alpha_2]s^3 + (4\alpha_1\alpha_2 + \omega_{20}^2)s^2 + 2\alpha_1\omega_{20}^2 s}{s^4 + 2[\alpha_1 + (n+1)\alpha_2]s^3 + (4\alpha_1\alpha_2 + \omega_{20}^2 + \omega_{10}^2)s^2 + (2\alpha_1\omega_{20}^2 + 2\alpha_2\omega_{10}^2)s + \omega_{10}^2\omega_{20}^2}. \tag{42}$$

The characteristic equation is now a quartic instead of a cubic. It may be factored into two quadratics in the manner proposed by Liu for fourth-degree equations (see Sec. 9·9).

$$(s^2 + 2\zeta_a\omega_{na}s + \omega_{na}^2)(s^2 + 2\zeta_b\omega_{nb}s + \omega_{nb}^2) = 0. \tag{43}$$

If the two quadratic factors[1] are multiplied out and coefficients equated to corresponding terms of the characteristic function (i.e., that in the denominator) of Eq. (42), four simultaneous equations are obtained:

$$\left.\begin{aligned} \zeta_a\omega_{na} + \zeta_b\omega_{nb} &= \alpha_1 + (n+1)\alpha_2, \\ 4\zeta_a\zeta_b\omega_{na}\omega_{nb} + \omega_{na}^2 + \omega_{nb}^2 &= 4\alpha_1\alpha_2 + \omega_{20}^2 + \omega_{10}^2, \\ \omega_{na}\omega_{nb}(\zeta_a\omega_{nb} + \zeta_b\omega_{na}) &= \alpha_1\omega_{20}^2 + \alpha_2\omega_{10}^2, \\ \omega_{na}^2\omega_{nb}^2 &= \omega_{10}^2\omega_{20}^2. \end{aligned}\right\} \tag{44}$$

The left-hand side contains parameters that indicate, in conjunction with nondimensional quadratic charts, the response properties of the system, while the right hand side contains the parameters α_1, α_2, ω_{10}^2, ω_{20}^2, and n which are defined in terms of the physical properties of the system and damper. One may thus use these equations in the same manner as described for the case of untuned damper to solve either the "response problem" or the "design problem." The only difference is that four simultaneous equations will be involved here instead of three. That is, given parameters α_1, α_2, ω_{10}^2, ω_{20}^2, and n, parameters derived from the physical parameters of the system, the set of four equations may be solved to give the four response parameters ζ_a, ζ_b, ω_{na}, ω_{nb} and thus the solution of the response problem. Or, given the parameter representing the load member of the original system, α_1, and given the response parameters ζ_a, ζ_b, ω_{na}, and ω_{nb}, then the set of four equations can be solved for the four unknowns α_2, ω_{10}^2, ω_{20}^2, and n which represent the design parameters of the tuned damper and the gain constant of the system.[2]

[1] The symbols of Brown and Hall are used here rather than the symbols used by Liu in Eq. (9·5).

[2] In the case of both the untuned and tuned dampers, the two types of problem considered may often be simplified by assuming that $f_1 = 0$, since f_1 will generally be small, and this approximation will therefore usually yield results sufficiently accurate.

The transient analysis of the damper has been considered in some detail, since it provides a good illustration of the method for treatment of higher-order error equations proposed in Sec. 9·9 as well as a satisfactory basis for utilization of a useful device. The feedback transfer functions of untuned and tuned dampers, given by Eqs. (36) and (45) respectively, might also be treated by the methods of frequency analysis as described in earlier sections, and the solution of response and design problems obtained in this way

$$\frac{\theta_o(s)}{E(s)} = \frac{\omega_{10}^2 s^2 + 2\alpha_2\omega_{10}^2 s + \omega_{10}^2\omega_{20}^2}{s^4 + 2[\alpha_1 + (n+1)\alpha_2]s^3 + (4\alpha_1\alpha_2 + \omega_{20}^2)s^2 + 2\alpha_1\omega_{20}^2 s}. \tag{45}$$

This latter approach becomes increasingly useful relative to the transient approach in case the servo system without damper contains additional energy storage or delay components. These would raise the order of the transfer function and make the transient analysis increasingly difficult. The treatment of these higher-order functions by the frequency method has already been discussed in considerable detail (Sec. 10·3), and its application to the transfer functions of the untuned and tuned dampers should offer no special difficulty.

Before closing this section, reference should be made to the analogy, already mentioned, that has been found to exist between the oscillation damper and special types of negative feedback network. Thus Harris[1] points out that a feedback network characterized by the transfer function

$$\theta_c(s) = \frac{-Rs^2}{1 + As} \tag{46}$$

is the control analogue of the type of damper designated above as an untuned damper. Here R is the ratio of damper inertia to output shaft inertia, and A is the ratio of damper inertia to viscous friction coefficient between the two inertias. In terms of the symbols used earlier in this section,

$$R = \frac{J_2}{J_1},$$

and

$$A = \frac{J_2}{f_2}.$$

Harris points out that this transfer function [Eq. (46)] is obtained electrically by differentiating a velocity signal in the type of network that also provides a time lag. Jofeh[2] also stresses this analogy, showing

[1] Harris, *op. cit.*, p. 72.

[2] Jofeh, *op. cit.*

explicitly the equivalence of the equations representing the two types of oscillation damper and two corresponding types of feedback network. The error equation for the untuned damper is shown to be equivalent to that of an electrical servo system in which a voltage proportional to the derivative of the output is fed back to the input of the amplifier through a condenser-resistance coupling. The error equation of the tuned damper is shown to be equivalent to that of an electrical system in which the tachometer voltage is applied to a circuit consisting of a condenser in series with a resistance shunted by an inductance, the voltage across the L-R combination being the input to the amplifier. He suggests that the analogues may be of practical value by making it easier to determine whether or not servo systems originally designed to work with negative feedback may be modified readily to work with oscillation dampers.

IMPROVEMENT OF ACCURACY

11·7. Correction of Transient Error.—The duration of the transient error may be given in terms of the time required for the error response curve to fall to less than a specified absolute value or to less than a specified percentage of its initial value. Or if the response is given in terms of output, a comparable definition is readily formulated in terms of the difference (or ratio) of the response at any instant and its final steady-state value. It is obvious that the specification of maximum allowable duration of the transient in such terms calls for a knowledge of the form and magnitude of the transient response, to be obtained either directly by way of transient analysis or experimental test or indirectly from frequency-response properties that may be correlated with the transient response. In case a comparison of the transient of a given system with specifications has been made on the basis of any one of these methods and its duration has been found excessive, its reduction then becomes necessary.

The *type* of corrective network appropriate for the reduction of the transient has been known for some time. It turns out to be the same type of device that was found effective in system stabilization, namely, one that takes the derivative of the error signal, i.e., one that introduces phase advance into the system. The association between derivative taking and reduction of transient time is fairly clear, since the taking of the derivative can be regarded as a way of anticipating future response and thus making possible more rapid error correction. The association between stability and small transient error is also readily made plausible by recalling that the system components that serve to delay the response of a network and thus increase the duration of the transient are thereby introducing time lags that predispose to instability. The equivalent of long time lags in the time response is phase lag in the frequency response, a phenomenon that, as we know, tends to produce system instability. Considered from the point of view of the Fourier analysis, a long lasting transient, or slow response, is equivalent to a spectrum in which the higher frequencies are attenuated. The frequency of the resonant peak of the $\theta_o(j\omega)/\theta_i(j\omega)$ response curve will, therefore, be low. The introduction of a phase-advance network will raise the frequency of the resonant peak, thus increasing the transmission of the higher frequencies. The equivalent time response to a step function will, therefore, be sharper.

The basic procedures for representing and designing phase-lead networks have already been considered in Sec. 11·4. It is unnecessary, therefore, to repeat this material here. What would be desirable at this point, however, would be a simple and direct method for relating the required reduction in the duration of the transient to the specification of the design parameters required in the phase advance network. Unfortunately, no such direct method seems to be available for any but the simplest systems. The most extensive attempt to develop design principles directed especially to transient reduction seems to be that of Hall[1] based on the transfer locus approach. The procedures for matching the corrective network to the system to be corrected may become quite involved; hence the original account should be consulted. Relevant principles formulated in terms of other points of view may be found in Draper and Bentley,[2] Brown,[3] Harris,[4] Ferrell,[5] and McColl.[6] In any case, the final test of whether or not the correct parameters have been chosen must be based on a new determination of the transient response. A mathematical cut-and-try process thus seems to some extent unavoidable.

11·8. Correction of Steady-state Errors.—As has been pointed out in previous sections, a number of different kinds of steady-state error may be defined, depending on the nature of the input test function. Procedures for compensating for steady-state errors of different order are similar. It will therefore be sufficient here to limit ourselves to a dis-

[1] A. C. Hall, *Analysis and Synthesis of Linear Servomechanisms*, Technology Press, Massachusetts Institute of Technology, 1943, "Theory of Phase Lead Controllers," Chap. 6, pp. 89–127.

[2] C. S. Draper and G. P. Bentley, "Design Factors Controlling the Dynamic Performance of Instruments," *Trans. ASME*, **62**, 428*ff*., July 1940.

[3] G. S. Brown, *op. cit.*, pp. 14, 39*ff*.

[4] H. Harris, "The Analysis and Design of Servomechanisms," NDRC Report, 1942, pp. 15, 17, 20, 24, 32–42, 5*ff*.

[5] E. B. Ferrell, "The Servo Problem as a Transmission Problem," *Proc. IRE*, **33**, November 1945, pp. 6*ff*. of reprint.

[6] L. A. McColl, *Servomechanisms*, Van Nostrand, New York, 1945, pp. 41, 104.

cussion of procedures for compensation of the velocity lag error. This type of steady-state error is the one which is generally of most importance in servo-system correction. It will be of interest to consider the nature of both the transient and frequency approaches to this problem.

Velocity-lag Error: Transient Approach.—The need for correcting servo systems for effects due to coulomb friction and constant torque loads applied to the output member has led to the development of the so-called *integral controller.* The need for integral controllers to counteract constant output torques was demonstrated by Minorsky[1] in 1922. Brown and Hall have recently analyzed the problem anew and reach the same conclusions.[2] They represent their analysis as an illustration of a general method for the design of controllers to meet specific conditions of excitation and performance. This method consists of first examining the error equation to see "what properties it must have to yield the solution desired" and second in "manipulating it mathematically to give it these properties." The manipulations consist in adjustment of the characteristics of the $C(s)$ operator in the error equation until this equation has the required properties. One may then attempt to synthesize physically a controller that can be represented by the expression finally derived for $C(s)$. In the present chapter, this general procedure was summarized in Sec. 11·4 in relation to the derivative error controller. In the case of the integral controller, the application of this general method indicates that the required form for the transfer function of the controller is

$$C(s) = \frac{T_c(s)}{E(s)} = k_0 + \frac{n}{s}, \tag{47}$$

where k_0 is the proportionality factor for proportional control and n is the proportionality factor for integral control. The integral equation for the controller corresponding to the transfer function of Eq. (47) is

$$T_c(t) = k_0E(t) + n\int E(t)\,dt. \tag{48}$$

If a controller of this type is introduced into a system with a characteristic equation of any given degree, the effect will be to increase the degree of the characteristic equation by one, as already shown in Sec. 9·8, for a system with a second-degree equation. Ways of solving such higher-degree equations have also been considered. Brown[3] and Brown

[1] N. Minorsky, "Directional Stability of Automatically Steered Bodies," *Jour. Soc. Naval Eng.*, **34,** No. 2, 280–309, May 1922.

[2] G. S. Brown and A. C. Hall, "Dynamic Behavior and Design of Servomechanisms," ASME Preprint, November 1945, pp. 14–20.

[3] Brown, *op. cit.*, pp. 26–36.

and Hall[1] carry out such solutions for a number of sample systems and discuss the effect of various relations among the parameters on the final form of the solution. What is of particular significance for us here is first the fact that a form for the controller transfer function can be found which will eliminate the steady-state velocity error due to various factors in the system and, second, the question of how integral controllers may be designed so that the over-all system will show the required performance properties. At this point it becomes desirable to shift to the frequency method of analysis, since such problems of design are more effectively carried out by this method.

Velocity-lag Errors: Frequency Approach.—In an earlier section, the relation between zero steady-state errors and the nature of the transfer locus as the frequency approaches zero was pointed out. Here, we may attempt to make use of these correlations in the design of servo systems with the required zero or minimal velocity-lag errors. Like principles are operative in adjusting systems for other types of zero steady-state errors, such as displacement errors and acceleration errors.

To show a zero velocity-lag error, the transfer function $\theta_0/E(j\omega)$ must have a second-order pole or greater at $\omega = 0$, and the locus must approach ∞ along the negative real axis. To achieve this condition, ways must be found of making the locus conform to this requirement. The obvious preliminary steps will be to plot the locus of the proposed systems, without any special adjustment or corrective device for velocity-lag control, and then determine the type of locus that must be multiplied with the locus of the given system to make it conform to the required condition of approach to ∞ along the negative real axis as ω approaches zero.

An example is given by Curve A of Fig. 10·20 (Sec. 10·5) regarded as representing a given servo system. This locus approaches ∞ by way of the negative imaginary axis. To meet the required conditions, the low-frequency portion must be rotated by $-90°$ so that it will run to ∞ along the negative real axis. Hence the given locus must be multiplied by a locus of the type represented by Curve B, one that will provide the required negative phase shift approximating 90° at the low end of the frequency scale, but not shifting enough to make the system unstable at the higher-frequency end. The product of Curve A by corrective locus B results in Curve C, which meets the required conditions for a zero velocity-lag error. If, now, the transfer function of Eq. (47), representing an integral controller, is plotted as a transfer locus, it is found to be identical with the type of locus shown by Curve B in Fig. 10·20. Thus two different approaches have led to the same result. A

[1] Brown and Hall, *op. cit.*, pp. 17–21.

controller with a locus of the nature of B is required for correcting a system with an initial velocity-lag error.

Can physical controllers be built that are represented by this type of transfer locus? Since the form of the locus implies infinite gain at zero frequency, it is necessary to use a regenerative amplifier.[1] A block diagram of such a controller is shown in Fig. 11·7. Hall has discussed the adjustment of the parameters of this network and demonstrated that

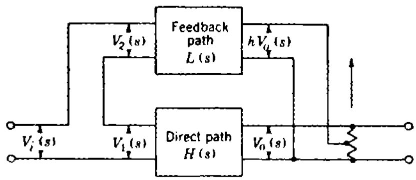


FIG. 11·7.—Block diagram of regenerative feedback amplifier.

for certain values of the network parameters, it will realize the transfer function given above for the integral controller. He points out, however, that one particular parameter h, in the circuit transfer function (49) below, must equal one, exactly, in order for this equation to reduce to the transfer function of the integral controller.

$$KG(s) = \frac{1 + RCs}{1 + RCs - h}. \tag{49}$$

Precise adjustment is often not possible, due to accidental variation of circuit parameters, so that this network usually approximates an integral controller but does not realize it exactly. If $h < 1$, the controller is called an undercompensating integral controller; if $h > 1$, it is called an overcompensating integral controller. The loci corresponding to these two types of controller are shown as the right and left half-circles of Fig. 11·8, and an ideal integral controller corresponding to $h = 1$, by the line. Under- and overcompensating controllers are designated as minimal velocity error systems, since the failure of the parameter h to equal one exactly gives rise to a system equation for which the velocity-lag errors can be made as small as desired but cannot be eliminated entirely. For detailed theoretical treatment of these systems, Hall's paper should be consulted.

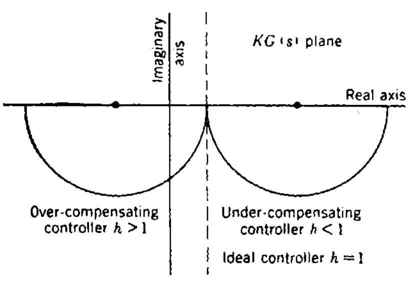


FIG. 11·8.—Transfer loci of various types of integral controllers.

[1] Hall, *op. cit.*, p. 66.

Before completing our brief survey of the integral controller, it should be noted that if a particular application permits the use of an undercompensating integral controller, the regenerative type of controller is not necessary, and this minimal velocity error controller can be physically realized by a filter (with the response characteristic of Fig. 11·9a) in cascade with an amplifier. This circuit is shown in Fig. 11·9b. Its locus and the general form of its transfer function are the same as that of the undercompensating regenerative integral controller.[1]

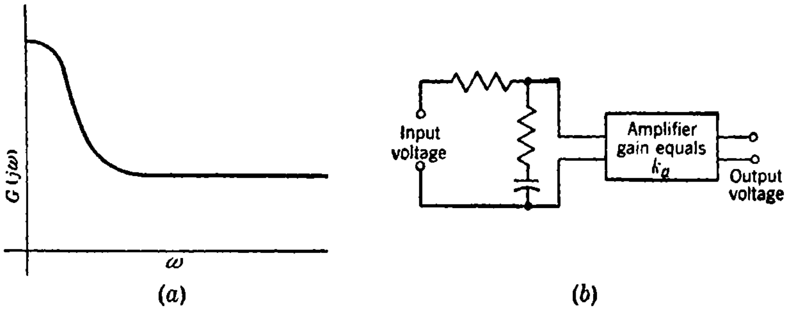


FIG. 11·9.—(a) Amplitude response of undercompensating integral controller which may be reproduced by passive filter circuit (reproduced from Hall, *op. cit.*, p. 86); (b) nonregenerative circuit having response curve shown in *a*.

SPECIAL PROBLEMS

11·9. Changes in Loop Gain.—In some servo systems, the special problem arises of stabilizing a system that is characterized by the equivalent of a variation in the loop gain. An example is the Arma-resolver servo system described in Sec. 14·3. In this system, the equivalent of a variation in loop gain is produced by the physical device used for error detection, an Arma resolver. This unit, a two-coil selsyn-type of device, can be regarded for present purposes as equivalent to an ideal error detector (used to determine the difference between input and output angles) in cascade with a converter (for converting angular error to an a-c voltage) in which the conversion factor varies in time under the influence of the input signal.[2] The nature of the application was such that the conversion factor changed during the course of a certain tactical operation. The problem arises then of how such a system may be stabilized, since standard servo theory assumes the constancy of all parameters.

For purposes of theoretical analysis, the system can be represented

[1] Hall, *op. cit.*, pp. 87*ff*.

[2] The error detector also contains a small nonlinear factor, since the error voltage is proportional to the sine of the angular error rather than to the angle itself. For small values of the angular error this difference may be neglected.

by the block diagram of Fig. 11·10. To consider questions of stability it is convenient to determine the feedback transfer function. We may assume that at a given instant, α, k_a, and k_m are constants. Then

$$KG(s) = \frac{\theta_o(s)}{E(s)} = \alpha k_a k_m H(s). \tag{50}$$

At a given instant, $\alpha k_a k_m$ will represent the servo gain K, and the system might be stabilized at any particular value of K on the basis of procedures already described. But at another instant, α and therefore K may be greater and the system may no longer be stable; in any case, the phase margin will have decreased. Conversely, α and therefore K may have become

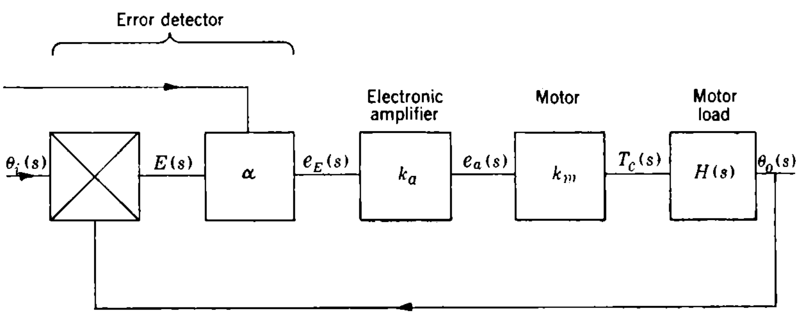


FIG. 11·10.—Block diagram of a servo system with variable loop gain. Symbols: α is error detector conversion factor $e_E(s)/E(s)$ in a-c volts per radian error, k_a is gain of electronic amplifier e_a/e_E in numerical units, k_m is motor conversion factor $T_c(s)/e_a(s)$ in torque units per volt applied, and $H(s)$ is transfer function of output member, in this case the motor load.

smaller. The transient response will consequently be slower and may not meet specifications. What means are available for avoiding these undesirable effects?

The problem can be simplified by first assuming that values for the amplifier gain and over-all gain have been selected to meet performance specifications over the full range and the system is to be stabilized for these conditions. Thus, for some representative value of α, a value of k_a may be selected to give an acceptably small static error, and a value for the over-all gain K to provide a sufficiently short transient. Under these conditions an appropriate stabilizing device can be designed, if needed, to provide a sufficient stability margin by procedures already described. We may thus think of a stabilizing unit introduced at some point in our block diagram of the system.

The problem is now reduced to finding ways of preventing instability when K increases owing to an increase in α and of avoiding the effects of a decrease in over-all gain K when α decreases. A review of the standard principles for system stabilization outlined in previous sections suggests the following possibilities:

1. Initially stabilizing the system for the maximum value of K expected to occur, as when α has its maximum value. The system will then be stable for all smaller values of α and K.[1]
2. Maintaining over-all gain K constant as α varies by automatically changing one of the other factors in the gain, such as k_a, or k_m, to compensate for changes in α. This scheme would thus require some method of automatic gain control.
3. Allowing the over-all gain K to vary with α, but automatically adjusting the stabilization network to produce a greater degree of correction as K increases and a reduced amount as K decreases.
4. Allowing the system, if nonlinear, to oscillate as α and K increase but setting the other factors in K so high that the frequency of oscillation will be high enough and the amplitude low enough so as not to exceed the allowable margin of error.

Method 1 may be considered the most conventional, and amounts to designing a nonadjustable system that has sufficient margins of safety to meet the worst conditions to be expected in practice. Methods 2 and 4 have both been tried and found satisfactory.[2] Method 3 has not, to the author's knowledge, been tried, but there seems no fundamental reason why it should not be feasible. If, for example, in the basic type of phase-advance circuit shown in Fig. 11·3 R_1 were decreased as α increased, such a decrease in R_1 would have the simultaneous effect both of reducing the over-all system gain and of advancing the phase, effects that would supplement each other in counteracting the effects of an increase in α. In any automatic adjustment scheme, it will, of course, be necessary to utilize some signal dependent on $\theta_i(t)$ or on α to control the adjustment. The selection of suitable circuits will depend on the nature of the physical components of the system in any given case.

11·10. Filtering and Data Smoothing in the Servo Loop.—The previous discussion of servo theory has been based on the assumption that the performance required of the output of a servo system is, within certain limits of tolerance, that of duplicating the form of the disturbance $f(t)$ imposed on the input (or some specific function of this input). Under operating conditions, however, the problem may be complicated by the fact that there may be superimposed on the correct signal $f(t)$

[1] This scheme would not work for a conditionally stable system, as is apparent from the discussion in Sec. 10·4.

[2] A. H. Fredrick at the Radiation Laboratory found that it was possible to use Method 4 effectively in the adjustment of nonlinear systems. Further details are given in Sec. 12·24.

an extraneous or undesired signal $g(t)$, either applied at the input or arising at some other point within the servo loop. This unwanted signal may be referred to as "noise." It may arise in a number of different ways: noise phenomena in vacuum-tube amplifiers, the chatter of gears, static friction effects in motor loads, interference effects in radar signals, and the like. If small enough, these noise effects may be neglected. At times, however, they may be great enough to exceed the acceptable tolerance of error in servo performance. The servo error must now be considered to be made up of two factors, a factor represented by poor following of the correct signal $f(t)$ and an additional factor due to good following of the noise signal $g(t)$. To eliminate this additional source of error, it is necessary to find some way of making the servo discriminate between the correct input function $f(t)$ and the incorrect one $g(t)$. The problem imposed on the servo is to follow $f(t)$ even though a function $f(t) + g(t)$ is imposed on the input. This is a problem of filter design. The servo system may thus be regarded as a special type of filter, and we may require the servo to satisfy certain filter specifications as well as the requirements of stability and accuracy considered in previous sections.

It will not be possible in the space available to go into this problem in detail. We may, however, review briefly the lines along which the problem may be attacked and indicate some of the sources where a more exhaustive treatment may be found.

Three types of approach have been followed or proposed. The more obvious and common approach has been based on the realization that a servo system (i.e., one containing mechanical elements) can be regarded as a low-pass filter. The amplitude transmission curve shows a certain range of transmission in the low-frequency region, followed by more or less sharp attenuation as the frequency increases. If the noise signal $g(t)$ is represented by a Fourier spectrum that lies *above* the frequency range which must be transmitted to realize the acceleration requirements of our servo system, then it should be possible to design the servo system so that its cutoff point lies below the frequency range of the noise signal. If such a design can be realized, the servo will be able to discriminate between the correct signal and the noise.

It is in this context that the problem of so-called data smoothing may be regarded as a special instance of a filter problem. In data-smoothing problems, we are interested in following the trend or average of a sequence of measurements or magnitudes rather than the instantaneous values, which will show more extreme fluctuations. Since these instantaneous fluctuations occur at a higher frequency than those of the averaged curve, the instantaneous fluctuations may be regarded as a high-frequency effect superimposed on the low-frequency variations present in the average curve. A servo system that is not capable of following the high-frequency fluctuations will thus have the effect of averaging out these fluctuations and hence of "smoothing the data."

In terms of the discussion of the preceding paragraphs, methods for filtering out the high-frequency noise sequence $g(t)$ or smoothing a given time sequence $f(t) + g(t)$ will be directed to reducing the higher-frequency response of the servo or, in filter terms, reducing the frequency of the cutoff point. There must be an increase in the inertia of the servo load or an equivalent change in the parameters of electrical networks. Such an increase in inertia will be accompanied, however, by an increase in the phase lags of the system and hence reduce the instability margin. It will therefore be necessary to introduce additional factors tending to increase damping factors or produce phase advance.

The approach just discussed might be designated, for brevity, as the "inertia method." It depends essentially on the common experience that mechanical or electrical inertia will have the effect of reducing the speed or frequency of response of a system and hence smooth out higher-frequency disturbances. It does not take into account the detailed nature of the two time functions to be separated or give any assurance that a given physical system designed on this basis is the best one for performing such a discrimination. A number of authors have recently proposed methods based on a more sophisticated mathematical background. A method proposed by Graham[1] for reducing the effects of signal noise uniformly distributed in frequency involves adjustment of the system so that it will show a transmission spectrum corresponding to the Fourier spectrum of the input signals expected to occur in practice. The method seems to involve some such sequence as the following. One first determines a time function $f(t)$ that can be regarded as representing the input signals to which the system will be required to respond. This time function is then represented as a frequency function by Fourier or Laplace transformation methods. Corrective networks are now incorporated in the servo system so the system as a whole will show frequency-transmission curves like that of the assumed signal. The system will thus pass signals of the specified form with a minimum of attenuation and all other signals, such as those involved in noise disturbances, with greater attenuation. Undesirable effects accompany the substitution of a limited pass band for a broader unselective pass band. These effects consist principally in the introduction of considerably increased time lags in the transient response. Just how great an increase in the duration of the transients can be tolerated for the sake of greater freedom from disturbing noise will depend on the specific application. A practical

[1] R. E. Graham, "Linear Servo Theory," *Bell System Tech. Jour.*, **25**, 616–651, October 1946.

difficulty in the application of this or any other method of analysis of the noise smoothing problem may consist in the difficulty of selecting standard time functions $f(t)$ and $g(t)$ as representations of the input and noise signals, respectively. The problem is of great importance in the design of automatic following radar equipment such as the SCR-584.

A more elaborate mathematical theory, originally proposed by Wiener[1] and supplemented somewhat by Daniells[2] and Levinson[3] seems to offer a more general basis for treatment of problems of this sort. The theory is not formulated specifically with reference to servomechanisms, but rather in relation to the more general problem of the design of filters and predictors for operating on a given time sequence $f(t)$ which may or may not be contaminated by a "noise signal" $g(t)$. It is pointed out, however, that servomechanisms fall within the scope of the theory. For the reader interested primarily in the possibilities for practical application of the theory, the report by Daniells and Levinson's second paper will probably be found most useful. These authors disclaim any attempt at mathematical rigor but aim rather to present the basic ideas in a way that will facilitate their practical use in engineering problems. The paper by Wiener and Levinson's first paper will be of more interest to the professional mathematician.

11·11. Nonlinearity in Servo Systems.—Practically all of the material available on servo theory is based on the assumption that the system being considered is linear, in the narrow sense already mentioned: that it can be represented by a linear differential equation with constant coefficients. These constant coefficients constitute the physical parameters of the system, such as J, f, r, and S; the assumption of linearity implies that such properties of the components may be regarded as constant, regardless of the magnitude of the forces applied or of the responses resulting. And yet, it is questionable whether any physical system is strictly linear, particularly when the events within the system exceed very small magnitudes. Why then has it been customary to analyze such systems by means of linear theory?

The reason, essentially, is that linear methods are available whereas

[1] N. Wiener, "The Extrapolation, Interpolation, and Smoothing of Stationary Time-series," *OSRD* Report No. 370, February 1942, pp. 176, available as U.S. Dept. Commerce Publication Board No. PB39700; abstracted *Bib. Sci. Ind. Reports, Dept. Commerce*, **3,** Nov. 8, 1946.

[2] P. J. Daniells, "Digest of the Manual on The Extrapolation, Interpolation and Smoothing of Stationary Time-series With Engineering Applications by Norbert Wiener," (British) OSRD ref. no. W-386-1, Mar. 25, 1943.

[3] N. Levinson, "An Expository Account of Wiener's Theory of Prediction," NDRC Applied Math. Panel Note No. 20, June 1945; "The Wiener RMS Error Criterion in Filter Design and Prediction," *Jour. Math. and Physics*, **25,** pp. 261–278, Jan. 1947.

nonlinear methods are not. There are no general methods for solution of nonlinear differential equations. Special methods that are available are relatively cumbersome and difficult to apply. Methods for analysis of linear systems are, on the contrary, usually comparatively easy to apply and turn out in practice to give results that are sufficiently accurate for engineering purposes. Predictions of system response are likely to fall within the limits of experimental errors of measurement; specifications of design parameters are likely to fall within the limits of precision of commercially available components. Final, small adjustments may then be made experimentally.

Yet cases do occur in which linear theory falls short—the system when built may show oscillations not predicted by theory, or it may not be possible to design a system that will meet the most stringent specifications on accuracy, for the assumption of linearity represents essentially an idealization of the physical system and therefore affords solutions that are only approximations. In cases in which the degree of nonlinearity of certain components is extreme or in which the response of a given component covers a considerable range, the approximation may be unsatisfactory. Nonlinear theory then becomes necessary. It would seem necessary, too, if we are to assess the proper scope of linear theory. For even in cases in which linear theory now seems adequate it hardly seems possible to determine the magnitude of the errors of approximation involved unless exact results may be obtained to check the approximations. Finally, development of a theory of nonlinear systems seems necessary in order to permit overt use of nonlinear components in achieving engineering objectives. Special phenomena arising in nonlinear systems have already led to some useful applications on an empirical basis. If a suitable theory for predicting the effects of nonlinear components were available, much more extensive use might be made of such components in servo design.

It will not be possible here to consider the problem in any detail. We shall limit ourselves to recording the chief units of the system in which nonlinearities of different kinds may occur, to pointing out the explicit attempts that have been made to deal with such nonlinear relations mathematically, and finally, to calling attention to the work going on in the more general field of nonlinear mechanics that provides a source of material for attack on the more special nonlinear problems of servomechanisms.

Following is a list of some of the units or parameters of a servo system in which nonlinear relations may occur:

1. *Nonlinearity in stiffness or torque/error ratio.*
    a. On-off control or relay servo systems. In such systems, torque is applied by the motor to the load only when the error exceeds a

given threshold. Then the maximum torque is applied and remains constant until the error falls below this or a related threshold. A graph of the torque-error relationship is not a straight line; the relationship is, therefore, designated as nonlinear.

b. *Torque saturation in motor.* Any motor has a maximum available torque, regardless of the magnitude of the error. Thus the output torque of the motor (as measured, e.g., with an opposing torque made large enough to stall the motor) may be proportional to the error up to a certain point and then approach a limiting value as the error is increased. The motor may therefore be considered a linear device until the error exceeds a certain critical value. For signals greater than this critical value, it will be operating as a nonlinear device.

2. *Nonlinearity in frictional or dissipation parameters.*

   a. *Coulomb friction.* Coulomb friction may be defined as a frictional force or torque the direction of which is always opposite to the direction of motion of the responding component. It thus depends on a property of the response; it is therefore to be considered nonlinear.

   b. *Nonlinear viscous friction.* A nonlinear viscous friction parameter may be indicated by a nonlinear torque-speed curve. As explained in Sec. 13·2, the decrease in torque as the angular velocity of the motor increases may be represented as due to an opposing frictional force that increases with the velocity. The slope of this line provides a measure of the magnitude of the viscous friction parameter f. If the slope is not constant, then f is not constant but is itself a function of the absolute velocity of the motor shaft. It must therefore be considered a nonlinear parameter, analogous in character to the nonlinear resistance of an electric circuit.

3. *Nonlinearity in inertia and other load parameters due to backlash.* Backlash in gear trains has the dynamic effect of a removal of the load on the driving shaft when this shaft reverses in direction or deaccelerates. At a given point, the parameters of the load are reduced to zero (in their effect on driving shaft) to resume their original magnitudes when the dead space between the gear teeth has been traversed. The magnitude of the load parameters thus depends on the nature of the response of the driven member in relation to the driving member and must consequently be considered nonlinear.

4. *Nonlinearity in amplifiers.* As is well known, the parameters of a vacuum tube μ, g_m, and r_p are not constants; yet they are com-

monly assumed to be constant in circuit computations, an approximation that is justifiable only while the tube is operating over a relatively narrow region and in the straight-line part of its characteristic curve. The magnitude of the tube parameters varies with the current through the tube and hence represents nonlinear relations between plate current and grid voltage or plate current and plate voltage. Such nonlinear effects in the amplifier of a servo system may show up in the behavior of the total system as the signals increase in magnitude.

5. *Nonlinearity in carrier conversion units.* Modulators and detectors may be used in servo systems to permit use of a-c amplifiers in place of d-c amplifiers. The modulator can be regarded as a component for converting a signal with a zero frequency carrier to one with a carrier of finite frequency. The detector later performs the reverse operation. The signal is represented by the form of the envelope. The conversion factor or ratio for either of these types of units is, as is well known, not a constant for all levels of signal but varies with the level. These units are thus inherently nonlinear.

To what extent have mathematical procedures been developed for analysis of servo systems involving such nonlinear relations? Unfortunately, the work so far reported is extremely meager. The number of papers published explicitly on nonlinearities in servo systems does not exceed six or seven. Hazen,[1] in his pioneer paper on the theory of servomechanisms, gives a graphical, step-by-step analysis of an on-off or relay type of servo system. Brown,[2] in his 1941 paper, gives a brief treatment of coulomb friction in servomechanisms. Minorsky,[3] in 1941, gave a brief review of nonlinear control problems, treated analytically. Harris,[4] in 1942, discussed the nature of some of the nonlinearities that may occur in servomechanisms and possible methods of treatment. Hurewicz and Nichols[5] in 1944 presented an analysis of servo systems with torque saturation. This account includes a brief treatment of mathematical procedures for dealing with nonlinear speed-torque curves.

[1] H. L. Hazen, "Theory of Servomechanisms," *Jour. Franklin Inst.*, **218**, No. 5, 287–313, November 1934.

[2] G. S. Brown, *op. cit.*, pp. 27*ff*.

[3] N. Minorsky, "Control Problems," *Jour. Franklin Inst.*, **232**, 53, 451–487, 519–551, December 1941.

[4] H. Harris, "The Analysis and Design of Servomechanisms," NDRC Report, 1942 pp, 60–64.

[5] W. Hurewicz and N. B. Nichols, "Servos with Torque Saturation," Part I, RL Report No. 555, May 1, 1944; W. Hurewicz, "Servos with Torque Saturation," Part II, RL Report No. 592, Sept. 28, 1944.

Finally, McColl,[1] in 1945, gave a treatment of the on-off servomechanism based on what has been called a topological method of plotting curves representing the differential equation of the system in phase space. We may perhaps add to the list a preliminary attempt by Nichols and Kreezer,[2] in 1946, to utilize the decibel-frequency approach in representing the effects of backlash in servomechanisms.

As examination of the references cited will show, nothing of the nature of a general approach to the handling of nonlinearities in servomechanisms is available. The work so far reported tends to be introductory in character, involving special methods for each of the special types of nonlinearity and special assumptions that tend to restrict the range of application of the results. It is apparent that only the barest beginning has been made in this field and that the bulk of the work remains for the future. On the other hand, it is conceivable that no single or general approach can be found owing to the large variety of ways in which a system may fail to meet the requirements for linearity.

In attempts to develop suitable methods, some help may be obtainable from the fields of nonlinear mechanics. This may be defined as the branch of mechanics that deals with the solution of the differential equations of nonlinear physical systems. It may be thus thought of as the more general field in which the theory of nonlinear servomechanisms would form a special part. This reference to the more general field does not imply that one may find there general methods for handling nonlinear differential equations. The methods reported in the framework of nonlinear mechanics tend to have the same special character mentioned above. In this broader field, however, one finds a much greater wealth of literature than represented by the relatively few papers cited above which deal explicitly with servomechanisms. One finds, too, an effort being made toward a more general approach, as illustrated perhaps by the work on topological methods, and by Kryloff and Bogoliuboff's method of successive approximations. In the greater wealth of material available, one may hope too to find physical systems treated that exhibit

[1] L. A. McColl, *Fundamental Theory of Servomechanisms*, Van Nostrand, New York, 1945, pp. 107–125.

[2] N. B. Nichols and G. L. Kreezer, "Analysis of Servosystems Containing Backlash," unpublished, 1946. The procedure consists in assuming a sinusoidal oscillation imposed on the input of the system and determining the amplitude and phase of the first Fourier component of the nonsinusoidal output response. By assuming that this Fourier component can be used to represent the output response, to a first approximation, the ratio of output to error signal may be dealt with as described in Sec. 10·3 for strictly linear systems. The error of approximation can be checked against a step-by-step solution of the nonlinear differential equations of the system. The use of the first Fourier component to represent the nonsinusoidal response is similar to the procedure employed by Kryloff and Bogoliuboff in their method of treating nonlinear systems.

nonlinearities analogous to those found in servomechanisms. Thus considerable work has been done on the solution of the second-order differential equations of a single-mesh electrical circuit containing a nonlinear resistance. Such a nonlinear resistance would seem to be directly analogous to a nonlinear viscous friction. There should consequently be a carry-over from one problem to the other. As a matter of fact, practically all the methods so far introduced for treatment of nonlinearities in servomechanisms are found applied in a more general setting, in the field of nonlinear mechanics. To aid in the acquisition of a background in this latter field, a selection is given below of references to a number of useful review papers, taken from a bibliography compiled by the author of recent publications in the field of nonlinear mechanics.

Kryloff, N., and N. Bogoliuboff, *Introduction to Nonlinear Mechanics* (translated by S. Lefschetz), Princeton University Press, Princeton, N. J., 1943.

Mandelstam, L., N. Papalexi, A. Andronow, S. Chaikin, and A. Witt, "Exposé des recherches recentes sur les oscillations non-lineaires," *Tech. Phys. U.S.S.R.*, **2**, 81–134 (1935).

Von Karman, T., "The Engineer Grapples with Non-Linear Problems," *Bul. Am. Math. Soc.*, **46**, pp. 615–683, August 1940. (Good bibliography).

Minorsky, N., *Introduction to Non-linear Mechanics*, Part 1, "Topological Methods," Report No. 534, December 1944; Part 2, "Analytical Methods of Non-linear Mechanics," Report No. 546, September 1945; Part 3, "Non-linear Resonance," Report No. 558, May 1946; Part 4, "Relaxation Oscillations," David Taylor Model Basin, 1947; Edwards, Ann Arbor, 1946.

Kellar, E. G., *Mathematics of Modern Engineering*, Chap. 3, "Nonlinearity in Engineering," Wiley, New York, 1942, pp. 201–304.

LeCorbeiller, P., *Les systemes autoentretenues et les oscillations de relaxation*, Libraire Scientifique, Hermann & Cie., Paris, 1937, pp. 46.

Preisman, A., *Graphical Constructions in Vacuum Tube Circuits*, McGraw Hill, New York, 1943.

Van der Pol, B., "The Nonlinear Theory of Electric Oscillations," *Proc. IRE*, **22**, No. 9, 1051–1086, September 1934.

11·12. Miscellaneous Accuracy Considerations. *Gear Ratios.*—An important consideration in the design of servos is the balancing of factors affecting the choice of gear ratios. It is desirable from the standpoint of smoothness and stability to have a large reduction between the motor and the output device. From the standpoint of speed of response, maximum speed, and maximum power transferred to load, relatively lower gear ratios are desirable. The exact requirements may be more exactly stated if position servos and velocity servos are discussed separately. For velocity servos, the lowest speed at which the motor will run smoothly as a servo motor divided by the desired minimum output speed sets a lower limit on n, where $1/n$ is the gear reduction. The highest speed at which the motor may be run with a reserve of torque for acceleration, divided by the desired maximum running speed of the output, sets an upper limit on n. If it is necessary also to slew with the motor, the top speed of the motor divided by the desired output slewing speed sets another (and usually lower) upper limit on n. In order to meet these conflicting requirements, it may frequently be necessary to use a low value of n to allow rapid slewing and to rely on tachometer (velocity) feedback and high amplifier gain to lower the minimum smooth speed of the motors. Minimum smooth motor speeds as low as 4 rpm can be obtained in this way. For integrator applications, electrical means for removing the errors caused by rough operation at low speeds may be used (*cf.* Chap. 4). These conflicting requirements also lead the designer to favor motors with high top speeds, provided the minimum smooth speed is not sacrificed.

In a position servo, the requirements are much the same. The lower limit on the gear ratio is the smallest increment that the motor may be made to turn, or the amplitude of its "jitter," divided by the maximum position error allowable from these causes. As above, the conditions for minimum and maximum smooth operating speeds and slewing speeds also apply.

A very important consideration when dealing with servos where rapid response is required is the maximum load acceleration α_L.[1] This is given (neglecting friction) by

$$\alpha_L = \frac{nT}{J_m n^2 + J_L}, \tag{51}$$

where $1/n$ = gear reduction from motor to output,
T = motor torque,
J_m = motor inertia, including gears on the motor shaft,
J_L = load inertia, measured at the load.

By differentiating with respect to n and setting equal to zero, there results

$$J_m = \frac{J_L}{n^2}. \tag{52}$$

This tells us that *for maximum acceleration of the load, the inertia of the load as seen at the motor shaft*, J_L/n^2 should be *matched to the motor inertia*. This is also the condition for maximum power transfer to the load and is analogous to impedance matching in electrical and acoustical problems. The maximum acceleration measured at the load is

$$\alpha_{L\max} = \frac{T}{2\sqrt{J_m J_L}}. \tag{53}$$

[1] Although derived independently by the author, this treatment has subsequently been found in part to be equivalent to that of a memorandum from ADRDE dated Sept. 9, 1942, entitled "A Note on Gear Ratio Value in Servo Systems." See also, G. S. Perkins, "Analysis of Servo Mechanisms for Instrument and Power Drives," *Product Eng.*, **17**, No. 4, 332, April 1946.

For maximum acceleration in the presence of load friction or load torque, the gear ratio n should be increased somewhat. In the usual instrument servo, n^2 is so large that the inertial load at the motor is essentially that of the motor armature and the first few gears J_m.

This treatment leads to a number of simple criteria useful in choosing motors for high acceleration servos. Where freedom to choose gear ratios exists, the best motor from the standpoint of acceleration at the load is that motor having the highest value of $T/\sqrt{J_m}$. The T/J_m ratio is frequently erroneously taken as a figure of merit. This ratio can be changed merely by adding a gear to the motor! However, if for any reason the gear ratio n between motor and load is restricted to a value large enough so that the motor inertia as seen at the load is *large* compared with the load inertia, then the criterion for maximum acceleration at the load is that T/J_m be maximum. Where n is such that the motor inertia as seen at the load is *small* compared with the load inertia, the criterion is maximum T. Intermediate cases may be evaluated by substituting values in Eq. (51).

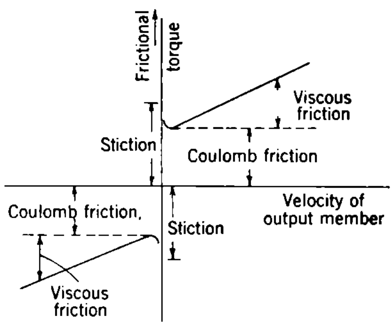


FIG. 11·11.—Diagrammatic representation of types of friction torque.

Friction.—It is necessary in most servos that operation at low speeds be smooth. A brief consideration of the factors determining whether or not smooth operation will prevail and what can be done to improve low-speed operation may therefore be of interest.[1]

We begin by defining terms. "Stiction" or "static friction" is the torque necessary just to initiate motion from rest. "Coulomb friction" or "dynamical friction" is that friction torque which opposes motion of the output and is independent of the output velocity. "Viscous friction" is a torque that opposes motion and is roughly proportional to the velocity of the output member. Figure 11·11 shows a diagrammatic representation of these types of friction. "Lumpy friction" is a torque varying with time and due to imperfections in gears, bearings, etc.

[1] Most of this discussion is based on the following British reports which have been generously made available: F. C. Williams, "Automatic Following Mirror Systems," TRE Report No. T1505, Part I, pp. 8–9; L. Jofeh, "The Effect of Stiction on Cyclic Control Systems," A. D. Cossor, Ltd, Research Department Report No. MR110; and N. R. Eyres, J. Howlett, J. Michel, and A. Porter, "A Theoretical Investigation of the Effect of Stiction and Coulomb Friction on the Minimum Smooth Running Speed of a Metadyne Controlled Motor as Incorporated in A.F.1," ADRDE Report No. 180.

The effects of these various types of friction in the case of proportional or proportional plus derivative servos may be summarized as follows. Stiction produces a "dead zone," that is, a range of input that produces no output motion. For a simple proportional servo, the total width of this dead zone is $2S/k$, where S is the stiction and k is the gain constant relating error and torque produced.

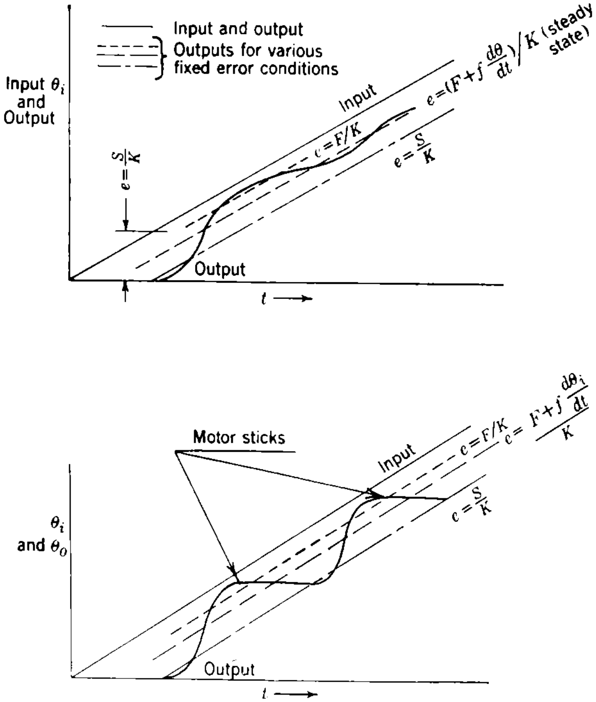


FIG. 11·12.—Servo operation showing effect of stiction.

Coulomb friction F is usually less than static friction. The difference between the two, $S - F$, may cause jerky operation at slow speeds. Once the output motion has stopped, error or misalignment e increases until the value S/k is reached, at which time the torque ke is equal to S. Motion then starts. Immediately the accelerating torque jumps to $(S - F)$, causing the output member to be accelerated in the direction that will reduce the error. As the velocity builds up and the error e is reduced, the accelerating torque $ke - F - f\,d\theta_o/dt$ falls toward zero and may reverse sign. In this expression, f is viscous friction, and $d\theta_o/dt$ is the output velocity. Where the values are such that the acceleration reverses sign, the resulting deceleration may cause the velocity of the output to fall to zero. Once this happens, stiction again rules, and no further motion will take place until the error again increases to S/k, giving jerky motion. "Lumpy" friction also produces jerky operation at low speeds, the process being very similar.

The literature on these effects is apt to be confusing. Both Williams[1] and Jofeh[2] make the mistake of assuming that motion will stop if e becomes less than F/k. For a system with inertia this is not the case. In Fig. 11·12a, reprinted from the ADRDE report,[3] a case is illustrated where e becomes less than F/k but stopping does not occur. Figure 11·12b shows a case where stopping does occur. The main conclusions of the ADRDE report, based on a differential analyzer investigation of a particular system, are

1. The critical value of $S - F$ is a function of K, f, and d_o/dt,
2. For given values of f and k, the minimum smooth running speed of the motor is directly proportional to $(S - F)$,
3. For given values of $(S - F)$ and $d\theta_o/dt$, the viscous friction f must increase with the "stiffness" k, and
4. For given values of f and $d\theta_o/dt$, k must be decreased as $(S - F)$ increases.

Williams[4] points out that for best low-speed operation it is desirable to achieve a large inertia/torque ratio, where the "inertia" may be either mechanical or electrical, and to incorporate a large amount of output velocity feedback coupled by a high-pass filter to retain d-c loop gain. This is helpful in reducing the effect of lumpy friction. He states as a convenient criterion of design that the feedback should be such that if the motor is suddenly stopped from the lowest running speed required, full torque is applied.

Backlash.—Because it is a nonlinear problem, backlash is difficult to deal with theoretically. This accounts for the great scarcity of theoretical information on the subject. A few simple generalities may, however, be given. The problem is of major importance only in position servos. In velocity servos, the load is usually large enough so that the gears run loaded, therefore eliminating the effects of backlash. At very low speeds, backlash does compete with stiction as a source of rough operation, although the effects are different, stiction causing jerky running and backlash causing oscillation or reduced stability.

In Fig. 11·13 is shown a schematic gearing diagram for a typical servo. This shows the three principal sources of backlash errors, namely,

[1] Williams, *op. cit.*

[2] Jofeh, *op. cit.*

[3] ADRDE Report No. 180.

[4] Williams, *op. cit.*

the gearing or couplings between A, input shaft, and data input device; B, output shaft and data output device; and C, motor and output shaft. The backlash requirement of each point must be considered separately.

Backlash at A causes errors in the output of exactly the value of the backlash, with appropriate factors for gear reductions applied. Backlash at this point has no effect on stability. Backlash at B causes errors in the output equal to the backlash as measured at the output shaft. In

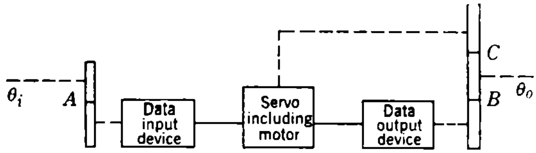


FIG. 11·13.—Schematic gearing diagram showing possible locations of backlash.

addition, this backlash contributes to instability, for it allows the motor velocity to build up before a reverse error signal can be applied. Backlash at C does not cause any errors in the output but contributes to instability. The output error due to the instability effects of the total backlash at B and C is reduced by increasing the fraction of this total occurring at C.

Based on the above considerations, standard practice is to use spring loaded or precision gears at A and B and reasonably good but not necessarily precision gearing at C.

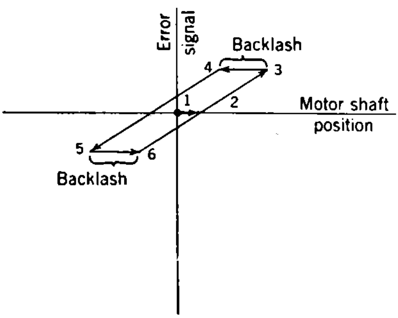


FIG. 11·14.—Error signal vs. motor position with backlash present.

A possible alternative to the arrangement illustrated is to gear or connect the data output device to the motor rather than to the output shaft. This is usually not preferred, since the output is then in error by the backlash errors of C, these usually being larger than those of B because heavier torques must be transmitted by C.

In d-c field-controlled motors, if field hysteresis errors are present, they act in a manner similar to backlash at point B. This fact should be kept in mind when designing or choosing a motor, particularly for position servo applications.

The following heuristic explanation of the effect of backlash may help the reader understand the instability caused by backlash. Figure 11·14 should be consulted. A fundamental explanation of the damping of a transient is that the energy dissipated by the system in a cycle of oscillation is greater than the energy gained in a cycle. The greater the ratio of the energy lost to the energy gained during a cycle the more rapidly will the transient damp out. The effect of backlash is to allow a torque ke_r (where e_r is the error in a transient at which motor reversal occurs) to act over an angle θ_B, the backlash, before the normal electrical damping means (operating on the error) come into effect. Thus, energy is added to the system each cycle, as compared with the same system without backlash. This extra energy added each cycle will reduce the ratio of energy lost to energy gained and hence reduce the damping or may even lead to oscillation. This explanation points the way toward a method of reducing the effects of backlash on instability, namely, use of damping means that operate directly on the motor shaft so that damping may be applied even during the backlash portion of the oscillation cycle. Among the methods included in this category are viscous damping, tachometer feedback (with either a-c or d-c coupling), and magnetic or viscous "flywheel" dampers.

CHAPTER 12

CHOICE AND DESIGN OF COMPONENTS

By I. A. Greenwood, Jr., S. B. Cohen,[1] D. MacRae, Jr., J. R. Rogers, W. D. Green, Jr., W. Goodell, Jr., and W. Roth

DATA INPUT AND OUTPUT DEVICES

12·1. Data Input and Output Devices: Introduction.—It is a fundamental feature of servomechanisms that a null or balance is sought by such systems. The principal purpose of data input and output devices is to furnish the necessary null or error signal to the amplifier and power-control devices or to furnish signals from which the error signal may be derived. Usually this involves a change of data representation as well as a discrimination, subtraction, or comparison of some kind. By way of illustration, a typical servo problem will be considered from the standpoint of these data devices. Input data are available as a certain representation, say shaft rotation, θ_1. Output data in the form of, say, shaft rotation are desired. In some manner, the output data must be compared with the input data, and an error signal derived representing the difference between the desired output (as a function of the input) and the actual output. This error signal may then be fed into the servoamplifier and used to actuate a motor or other power device in such a sense as to cause the error signal to be reduced by motion of the output. The first and second synchros in Fig. 8·1 are typical data input and output devices, respectively.

Data input and output devices may be conveniently divided into two general types, depending on the manner in which error signals are derived. The first type supplies an error signal directly. The synchro data system of Fig. 8·1 is such a system, for the signal from the second synchro directly represents the error in the position of the output. The second type yields two separate signals that must be compared (discriminated) in some way in order to obtain an error signal.

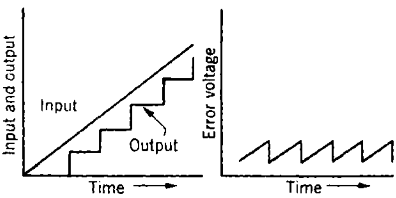
It is not always necessary that both a data input and a data output device be present, since sometimes input or output data may be in a form suitable for direct use in the control circuits without further change of representation. For example, in supplying an error signal to a servo motor positioning a shaft such that its rotation is proportional to a

[1] Sections 12·1 to 12·8, inclusive, are by I. A. Greenwood, Jr., and S. B. Cohen.

voltage derived from a preceding computer element, this computer voltage may be used directly in the same way as a voltage derived from an output data device such as a potentiometer, no input data device being required.

In choosing an input or output device, one should apply the list of design factors of Chap. 19. Factors that are particularly important with regard to input and output devices include resolution, speed of rotation, mechanical loading, electrical loading, accuracy, a-c or d-c operation, frequency of alternating current, life, interaction of units operated in parallel, size, weight, complexity of associated equipment, and range of operation.

When an input or output device with poor *resolution*, such as a coarse potentiometer, is used in a servo following a smooth, small velocity input, an interesting type of jerky operation may occur. This is illustrated in Fig. 12·1. The lack of smoothness, measured quantitatively by the percentage variation in instantaneous velocity, has been shown to be inversely proportional to the mean velocity of rotation and the number of discrete steps per degree. Increasing the damping ratio reduces the jerkiness somewhat. For constant damping ratio and constant number of steps per degree, the speed at which a given percentage variation in velocity obtains is directly proportional to the undamped natural frequency of the servomechanism. While jerky operation due to this cause is to be guarded against, use of ordinary high-resolution potentiometers is usually sufficient to make the effect negligible. Errors of approximately the magnitude of the resolution will, of course, be introduced in positioning servos when devices such as potentiometers are used as data input or output devices.



(a) (b)

FIG. 12·1.—Characteristics of jerky motion caused by insufficient resolution.

Speed of rotation affects life and quality of moving contacts. *Mechanical loading* may be an important factor in the choice of a data input device. For very lowest loads, instrument Magnesyns, Microsyns, Telegons, Microtorque potentiometers, and other similar devices are to be recommended. These are discussed below. *Accuracy* is usually expressed as an angular limit error for synchros and other similar cyclic devices and as a percentage independent linearity limit error[1] for potentiometers and other similar devices.

[1] The term "independent linearity error" refers to deviations from the best fit straight line for the plot of output voltage or percentage of total resistance vs. shaft rotation, both slope and zero being adjusted in the fitting process.

Interaction of parallel units[1] is of importance, particularly in the case of synchros, where conditions may lead to anything from negligible errors to prohibitive errors. The principal factors are electrical loading and rotor design. This subject will be discussed below in more detail.

12·2. Synchros and Related Devices.—*Synchros* are among the most common and most important data input and output devices. They may be characterized by low mechanical loads, long life, suitability for parallel operation if not excessively loaded, moderate to high accuracies (1.5° to 0.1°), continuous rotation, a-c operation, and in some sizes light weight. Their theory of operation has been discussed in Vol. 17 of this series. As has been pointed out, synchros provide direct error signal, no comparison of signals being necessary.

In servo systems, synchros are used as (1) generators, (2) control transformers, and (3) differentials. Their use as synchro motors will not be treated here. Used as control transformers in a servo system, synchros are considerably more accurate than as synchro motors. Several synchro control transformers may be used in parallel driven by a single synchro generator. The errors of such a system are minimized by (1) light electrical loads on the control transformer rotors, (2) use of distributed-winding type rotors, and (3) operation of all rotors at or near null. If the latter condition obtains, larger electrical loads are permissible.

Navy type synchros[2] are designated by a number 1, 3, 5, 6, 7, or 8 which indicates the size and is roughly equal to the weight in pounds. Size 5 is the most common size 60-cycle synchro and is the largest-size synchro motor. Errors for this series may run from 1.5° for Size 1 to 0.5° for Size 5. A particularly accurate, small, and light series of 400-cycle synchros for airborne use is sold under the trade name of AY-100 series Autosyns.[3] Such a synchro is shown in Fig. 5·12. Accuracies of these synchros may run from $\frac{1}{4}°$ to $\frac{1}{10}°$.

The *Magnesyn*[4] is another type of data input device. In servos it is usually used with a synchro control transformer. Because it has no brushes or slip rings the Magnesyn is particularly useful for applications where the mechanical load introduced by the data device is required to be extremely small. Its use is illustrated in Fig. 12·2 and also in Fig. 12·64, Sec. 12·23.

The Magnesyn consists of a laminated core on which a toroidal coil is

[1] See Vol. 18 of this series.

[2] Army type synchros, although similar in characteristics, were not classified by any such simple system.

[3] Manufactured by Eclipse-Pioneer Division of Bendix Aviation Corp., Teterboro, N. J.

[4] R. S. Childs, "Magnesyn Remote Indication," *Trans. AIEE*, **63**, 679–83, September 1944. "Magnesyn" is a trade name of the Eclipse-Pioneer Division of Bendix Aviation Corp.

wound. The rotor is a small permanent magnet. The coil is excited by 26 volts of 400-cycle power. The constants of the system are so chosen that the core will be completely saturated at each peak of the excitation current. Owing to the unsymmetrical saturation of the core when it is magnetically biased by the flux of the rotor, alternating voltages of second and higher even harmonic frequencies are set up at the three take-off points of the coil. The relative values of the voltages at these points depend on the position of the rotor. Leads from the three points are connected to the three stator windings of an Autosyn through a special transformer canceling voltage components at the

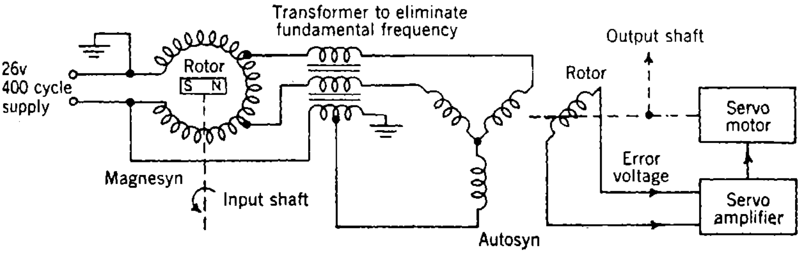


FIG. 12·2.—Use of Magnesyn in servo loop.

exciting frequency. The Autosyn is operated as a control synchro in the usual manner. The phase detection of the error signal from the Autosyn is, of course, done at the second harmonic frequency. Torque required to turn a Magnesyn varies widely but is at least an order of magnitude smaller than the 0.04 in.-oz of the Autosyn-type synchros. Accuracies of approximately $\pm\frac{1}{4}°$ are obtained.

A Magnesyn has been used to advantage as the data input device in a servo follow-up actuated by the low-torque output shaft of an air-speed meter and has also been used frequently in other similar applications. A considerable number of Magnesyns were used in conjunction with Bendix Gyro Fluxgate compasses as input elements in follow-up systems using compass data, an important example being in the north-stabilization of the AN/APS-15 radar plan position indicator.[1]

Another data input or output device of interest is the *Microsyn*.[2] A Microsyn is an electromagnetic device which has a specially shaped two-pole, soft iron rotor and a four-pole stator. Coils are wound on each stator pole. The characteristics of the unit are determined by the type of windings and connections used and by the shape of the rotor. The basic Microsyn can be so designed and connected as to provide (1) an a-c voltage varying linearly in amplitude with the product of the exciting

[1] See Vol. 22 for details of PPI north stabilization.

[2] The Microsyn was developed at the MIT Instrumentation Laboratory but is not yet available commercially.

current and the angular rotation of the rotor from a null or initial position, (2) a torque output that is proportional to the square of a d-c input current or to the product of the currents in two exciting windings but is essentially constant over a reasonably large angle of rotation of the rotor, or (3) a torque output that is proportional to angular rotation and varies with the square of the exciting current. The connection shown in Fig. 12·3 corresponds to Case 1. The output is alternating current, and its relative phase changes by 180° as the rotor goes through the null position.

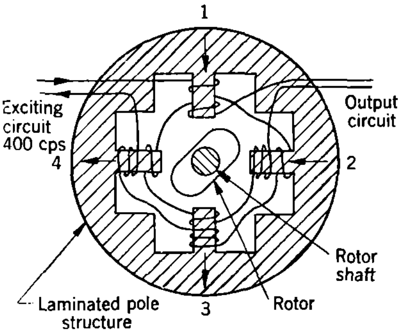


FIG. 12·3.—Schematic diagram of Microsyn.

As used for a data input or output device for servo use, the Microsyn is not itself positioned for minimum voltage, but rather the output unit positioned to give a voltage of opposite phase and of magnitude equal to that of the voltage from the input unit. The algebraic sum of the two voltages is then near zero and constitutes the error voltage input to the servo amplifier. A Microsyn is normally used only over a limited angular range.

Because each Microsyn is separately excited, it is possible by the use of high-impedance comparison devices to position several units so as to follow an input unit or to make a series of units position in cascade. Microsyns require controlled exciting current for some applications; for other applications it is sufficient to pass the same exciting current through both the input and output units.

An important advantage of the Microsyn lies in the fact that, like the Magnesyn, it requires no slip rings for the rotor and therefore its torque load is only that of the shaft bearings, which may be made very small. One small Microsyn unit has an outside diameter of about 2 in. An amplitude vs. shaft rotation plot of the output is linear to about $\pm 0.08°$, and two units will track together in a null-balancing circuit to within $\pm 0.06°$ for angles up to $\pm 7\frac{1}{2}°$. Null operation with somewhat reduced accuracy is possible out to $12\frac{1}{2}°$ either side of the zero position. Normally the exciting windings are low impedance, of the order of 10 ohms at 400 cycles, and the output windings have an effective impedance of the order of 2500 ohms.

The *Telegon*[1] is another member of the synchro family. By virtue of a unique type of construction, it, like the Magnesyn and Microsyn, has

[1] Telegon is a trade name of the Kollsman Instrument Co.

no moving contact. This results in a minimum of radio and other electrical interference, long life, and extremely light torque load.

A Telegon includes two stator coils, as in an ordinary synchro. The rotor consists of a shaft turning in jewel bearings and carrying two soft iron vanes. The rotor is magnetically excited by an a-c coil concentric with its shaft. The vanes serve to distort the field produced by the rotor excitation coil such that a maximum flux linkage with the stator structure occurs in a plane containing the vane axis and the shaft. This plane, of course, rotates as the rotor is turned, thereby causing induced voltages in the two stator coils to vary nearly as the sine and cosine of the rotor angle. Since Telegons are normally used in pairs, it is not essential that the stator voltages vary exactly sinusoidally with rotor rotation but merely that the variation be identical for units used together.

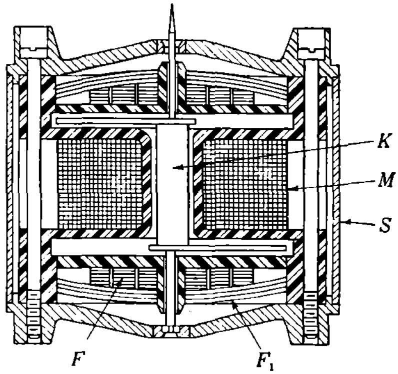


FIG. 12·4.—Cross section of Telegon. (*Courtesy of Kollsman Instrument Co.*)

Telegons are available for 400-cycle, 110- and 26-volt excitation and for 700-cycle, 85-volt excitation. Their average weight is approximately 4 oz.

The physical arrangement of parts is shown in Fig. 12·4. Winding M is the concentric rotor-exciting coil, while windings F and F_1 are the stator coils. The axes of the three windings are mutually perpendicular. Part K is the rotor with its soft iron vanes, the remainder of the flux path being completed by a soft iron shell S, which also serves as a magnetic shield.

12·3. Potentiometers.—Although having some disadvantages as compared with the devices discussed above, potentiometers as input or output devices are often more practical because of their small size, small weight, low cost, ready availability, low power requirements, possibilities for d-c operation, and adaptability to specific systems. Disadvantages of potentiometers as data input and output devices sometimes encountered are mechanical loading, electrical loading, speed limitations, short life, resolution, and insufficient accuracy. This list of possible shortcomings of potentiometers should not be interpreted as a blanket indictment of all potentiometers, for there are a number that do not fall down on any of these counts, and there are many applications where some of these characteristics are of no importance. Potentiometers are treated in detail in Vol. 17 of this series. This discussion will be limited to those characteristics particularly important when potentiometers are used as data input and output devices.

A potentiometer is generally lighter and smaller than a 60-cps synchro capable of doing the same job but is about equivalent in size and weight to a number of 400 cps devices such as Autosyns, Telegons, etc. In servo applications, potentiometers are usually, but not always, operated as high-resistance (say 20,000 ohms) potential dividers and hence require little power for energizing. In comparison, synchros draw considerably more total power, although this may be directly from the a-c line and therefore not particularly difficult to obtain:

Potentiometers are suitable for use with both direct and a-f alternating current. They are particularly flexible, since they can be used in a variety of electrical combinations as either variable resistors or potential dividers. They are particularly useful as output devices, allowing servos to follow input voltages of any desired nature. In general, potentiometers require more torque than synchros, 1 to 3 oz-in. being typical. Special potentiometers have been built with very low torque however, an example being the Microtorque potentiometer[1] having a load of only 0.003 oz-in.

A frequent restriction on the use of a potentiometer may be the requirement that it feed a low-impedance load. Normally, this introduces a loading error which may be quite serious. However, if one potentiometer is matched against another, the electrical loading of a balanced servo input circuit connected between the two arms will not cause any inaccuracy in the null point. The speeds at which potentiometers may be used vary widely; in general the maximum speeds are appreciably less than the maximum speeds for synchros or condensers. The speed problem is discussed in Vol. 17. Intimately related to speed is the useful life of a potentiometer. Life is an important consideration in the choice of an input or output device. From this standpoint, synchros, condensers, and other types of slip ring or no-moving-contact devices are preferred to potentiometers, particularly for the higher-speed applications. Considerable progress has been made in the design and construction of potentiometers for long life; if a potentiometer is properly made and is not operated at excessive speeds, a useful life of the order of millions of cycles is possible.

The accuracy and resolution of a potentiometer is also a very important consideration as regards its use as a data input or output device. Where desired, the resolution and accuracy of a potentiometer may be made very high; for example, in the 10-turn Helipot[2] or Micropot[3] type

[1] See Vol. 17. Made by G. M. Giannini and Co., 161 East California St., Pasadena, Calif.

[2] Trade name of the Helipot Corporation, 1011 Mission St., South Pasadena, Calif.

[3] Trade name of Thomas B. Gibbs and Co., Delevan, Wis.

of construction, accuracies of 0.02 to 0.10 per cent independent linearity limit error and resolutions of one turn of wire equal to 0.01 per cent of full scale are available in 20,000-ohm potentiometers. In the RL-274 and RL-270[1] single-turn potentiometers, accuracies of approximately ±0.07 per cent and resolutions of 0.03 per cent of full scale are obtained.

Accurate, low-resistance potentiometers and variable resistors were designed and manufactured in small quantities by the Minneapolis-Honeywell Regulator Company.

12·4. Condensers.—Condensers are useful for data transmission and particularly as data input devices because of their low torque loads and the high speeds at which they may be used. Condensers for high-speed applications are generally built without moving contacts, and therefore only the frictional torque of bearings need be overcome in rotating the shafts. The low torque characteristics of a condenser data input device have been used successfully and to great advantage in the MIT differential analyzer.[2] It is extremely important in the operation of this machine that the data take-off which transmits the motion of the output disk of each integrator impose a minimum mechanical load on the disk. In this application the condenser is of the rotating dielectric type. A pair of condensers whose capacitances are varied as the sine and cosine of the input shaft rotation constitutes a transmitting or receiving element of this data-transmission system.

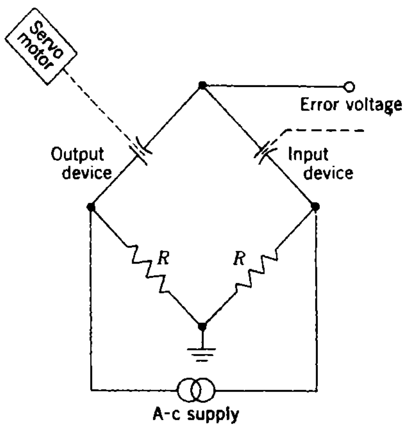


FIG. 12·5.—Use of condensers as data input and output devices.

Figure 12·5 shows another common application of condensers as data-transmission elements, in which a bridge circuit is used to obtain an error voltage. This circuit, of course, has the restriction that only the range of data covered by the limited range of operation of the condenser can be covered, cyclic operation being impractical. As in most a-c bridge circuits, it is necessary to take precautions about ground capacities, etc.

Another application of a rotating condenser is found in the Teleplotter equipment.[3] In this application a beat-frequency type of circuit

[1] Designed by Radiation Laboratory and manufactured by the Fairchild Camera and Instrument Co. and the Muter Corp.

[2] V. Bush and S. Caldwell, "A New Type of Differential Analyzer," *Jour. Franklin Inst.*, **240**, 4, October 1945.

[3] Brown Instrument Co., 4428 Wayne Ave., Philadelphia 44, Pa.

is employed, in which each condenser input and output is part of the tuned circuit in a very stable oscillator. When the input and output positions are identical, other constants in the circuit are so adjusted that identical frequencies in each oscillator circuit are produced. A selective filter produces an error voltage proportional to the beat frequency that is present if the condensers are not in corresponding positions, and this error voltage is used in the usual manner to position one of the condensers so as to reduce the error signal. This method is advantageous for radio or wire transmission of data to remote points, the frequency of the input condenser oscillator being the data representation in this case.

The accuracy of phase-shifting condensers is an important consideration in their choices as data input and output devices. The best condensers available to date have peak errors of approximately 1° and are available[1] in three- and four-plate styles for use with three- and two-phase oscillators, respectively. For further details see Vol. 17, Chap. 9.

12·5. Miscellaneous Data Input Devices.—There are an endless variety of input devices that may be used to transform physical measurements into electrical or mechanical forms suitable for use in instrument servos. Detailed discussions of this subject may be found in the literature of process control.[2]

12·6. Time-derivative Data Input and Output Devices.—The use of time-derivative reference devices in velocity servos is discussed in detail in Chaps. 4 and 14. In choosing an output device for a velocity servo the usual list of design considerations of Chap. 19 should be kept in mind. However, the following characteristics are of particular importance: weight, accuracy, life, voltage, scale factor, a-c vs. d-c operation, variable vs. constant frequency if alternating current, and means for controlling the voltage scale factor. For d-c circuits several d-c generator type tachometers are available or in development.[3] Condenser tachometers have also proved themselves valuable for some applications. A condenser tachometer has the advantage that its output may be the product of an input d-c voltage and the speed of rotation of its shaft. Alternating current tachometers are generally desired with constant frequencies. These may be obtained by exciting one winding of a two-phase induction motor type device with a constant-frequency alternating current, obtaining a voltage proportional to speed of rotation

[1] P. J. Nilsen Co., Chicago, Ill.

[2] See, for example, E. S. Smith, *Automatic Control Engineering*, McGraw-Hill, New York, 1944; Eckman, *Principles of Industrial Process Control*, Wiley, New York, 1945; and R. R. Batcher and W. Moulic, *Electronic Control Handbook*, Caldwell Clements, 1947.

[3] See Vol. 17 of this series.

from the other winding. The Kollsman[1] and Arma[2] drag-cup generators are examples of the better devices available for this use.

Alternating current tachometers generating variable-frequency alternating current may also be used, generally in conjunction with an input device or signal characterized by a frequency that is to be matched. This scheme is particularly useful where it is desired to maintain accurately a fixed speed of rotation. An example of such a circuit is a Bell Telephone Laboratory crystal-controlled motor. Variable frequency a-c tachometers might also be used in conjunction with frequency-measuring circuits, but this is seldom done. In some applications it is necessary to have the velocity servo operate for only a short period of time; in such cases it may be advantageous to use a potentiometer with direct current across it as the output device, differentiating the output voltage from the arm of the potentiometer by means of a network to obtain a voltage proportional to velocity.[3]

Time reference devices may be used in positioning servos as stabilizing devices, since a voltage proportional to speed added to the normal error signal in a linear system is equivalent to viscous damping applied to the motor. This stabilizing technique is widely used, both with d-c and a-c velocity references in d-c and a-c servos, respectively. For this application the accuracy of the velocity reference device is relatively unimportant. When d-c velocity feedback is used for stabilizing purposes, a large condenser is often inserted in series with this feedback to avoid the increase in velocity error due to the viscous damping effect while still retaining the stabilizing action at the higher frequencies where such action is needed. The stabilization of servos by means of velocity feedback is discussed in Chap. 11.

Occasionally a voltage proportional to the second derivative with respect to time of the output or input is desired. Possible but not common applications might be for extra stabilization of velocity servos or in servos where output data proportional to the second derivative or integral with respect to time of input data is desired. One method of obtaining such voltages is to differentiate the voltage obtained by use of a velocity device; however, this is usually not very satisfactory because of its tendency to exaggerate commutation noise. The subject of higher time derivative elements is discussed in Chap. 4. Mechanical devices taking advantage of the relation $F = ma$ are sometimes used as acceleration-measuring devices.[4]

[1] Models 915-B-04602 and 863-04302, Kollsman Instrument Division of Square D Co., 80–08 45th Ave., Elmhurst, N.Y.

[2] Arma Corp., 254 36th St., Brooklyn, N.Y.

[3] See Chap. 4.

[4] See, for example, C. S. Draper and W. Wrigley, "An Instrument for Measuring Low Frequency Acceleration in Flight," *J. Aero Sci.*, **7**, 388–401, July 1940.

12·7. Nonlinear Positioning.—It is often desired that the output motion of a servo system be some nonlinear function of the input data; this situation is frequently encountered in computers. In a servo loop the nonlinearity may be introduced in either the input or the output device. The input or output device may be nonlinear or may be nonlinearly connected to the input or output data. An example of the former method might be a nonlinear potentiometer; an example of the latter method might be a linear potentiometer connected to the output shaft by means of nonlinear gears. The techniques for producing such nonlinear transformations of data have been discussed elsewhere;[1] however, a few general considerations that affect over-all servo loop performance are of interest here.

If the nonlinearity is introduced in the input device or in the connection between the input data and the input device, then the servo system is not subject to change of gain and nonlinearity in the servo loop. If, however, the nonlinearity is introduced in the output device or in the connection between the output data and the output device, then this nonlinearity becomes a part of the servo loop and affects its analysis from the standpoint of change of gain, nonlinearity of loop, or both.[2] Thus, one important conclusion may be immediately drawn: If a choice is possible, *the nonlinearity should be introduced at the input element or between the output data device and the load; i.e., outside the feedback loop.*

In Fig. 12·6*a* a linear potentiometer is used as an input device and a resolver as an output device, yielding a servo loop in which the sine of the output is proportional to the input. In Fig. 12·6*b* the functions of the input and output devices are reversed, resulting in a system whose output is proportional to the sine of the input. In the first system, the loop gain varies; in the second system, the loop gain is constant.

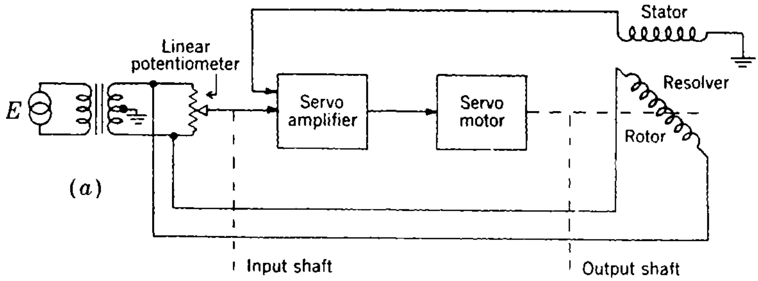
12·8. Single-speed vs. Multispeed Data Transmission.[3]—In systems where the accuracy of data transmission required is greater than can be conveniently obtained with synchros operating as shown in Fig. 8·1, it is common practice to use a second synchro channel operating at a higher speed relative to the first channel, such as 36 speed. The error signals are then taken from the high-speed synchro channel. This method has two advantages: It reduces the effect of synchro inaccuracy inversely as the speed of rotation is multiplied and results in an increase of servo gain proportional to the multiplication of speed. Of course, means must be provided for controlling from the output of the 1-speed

[1] See Chap. 5.

[2] *Cf.* Sec. 11·11 for a discussion of the effects of nonlinearity of different types on servo performance and analysis.

[3] *Cf.* Sec. 2·11.

synchro channel until the right revolution of the higher-speed synchro has been reached, i.e., the systems must be synchronized. In some systems, such as the British remote-reading aircraft compass, the function of the 1-speed element is fulfilled by other means. In the case of the remote-reading compass this is by voice communication over an interphone system. The major disadvantage of the multispeed data-transmission system is, of course, the additional complexity and equipment involved.



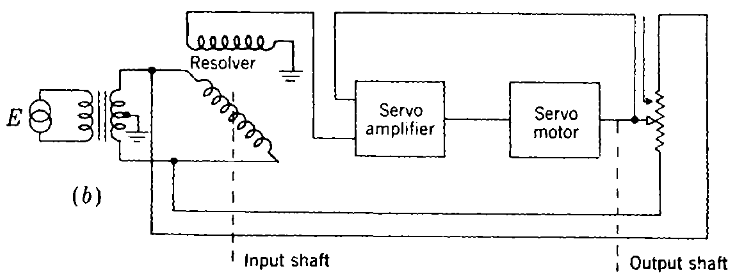


FIG. 12·6.—Examples of variable-gain type nonlinear servo loops.

The choice of the exact speed of the high-speed element depends, among other things, on the relation between the accuracy of the high-speed element and the over-all accuracy required, and on the relation between the velocity of the 1-speed element and the maximum speed allowable for the high-speed element. As an example of this limitation, synchros are usually rated at 1000 rpm, except for Sizes 1 and 5F, which are limited to 400 rpm.

By way of illustration, consider a system using synchros with accuracies of $\pm\frac{1}{2}°$. If in addition to a 1-speed synchro channel, another synchro channel were added at 36-speed, the error of following (measured at the 1-speed shaft) due to the synchro errors would be reduced to only 0.014°, and the gain of the servo system would be increased by a factor of 36. Figure 12·7 is a block diagram of such a 1- and 36-speed servo system. Block A in this figure conta ns the data input devices, in this case 1- and 36-speed synchros. These drive 1- and 36-speed synchro control transformers in block B. The control transformers are lined up so that the electrical zero position of the 1-speed device coincides with an electrical zero position of the 36-speed device. As indicated, the rotor outputs are fed to a single-pole double-throw relay. The normally open position of the relay connects the 36-speed information to the

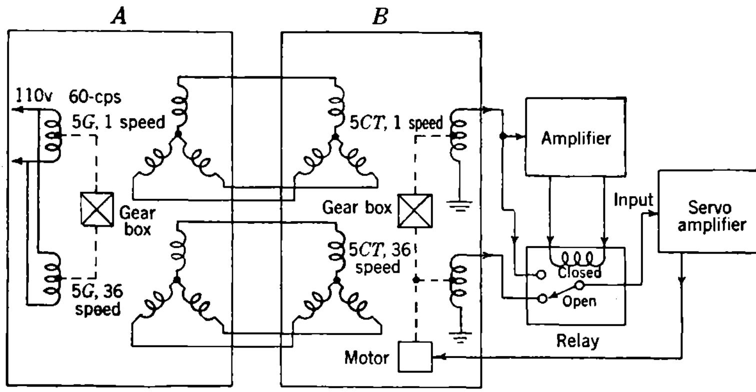


FIG. 12·7.—Servo with 1- and 36-speed data transmission.

servoamplifier. The 1-speed error voltage is fed to an amplifier that closes the relay only if the error voltage from the 1-speed control transformer corresponds to more than a 5° angular position error. If more than a 5° error is present, the servoamplifier is operated from the 1-speed error voltage. As soon as the servomotor has decreased the error of following to less than 5°, the relay will open and the 36-speed synchro control transformer will again take control of the servoamplifier. Thus, the servo will always operate in the correct sector with no ambiguity. In Fig. 12·8 is shown an amplifier circuit suitable for this type of servo operation. As the 1-speed error voltage increases, a value of grid voltage of the first tube is reached at which the plate current closes the relay, connecting the 1-speed error voltage to the input of the servoamplifier. A number of alternate schemes for this type of operation of multispeed systems have been developed. Some of the simpler circuits use series or parallel addition of the two synchro rotor voltages with contact rectifiers, diodes, or neon lamps to limit the voltage from the high-speed data synchro to a value which insures that the 1-speed synchro will assume control for large errors.

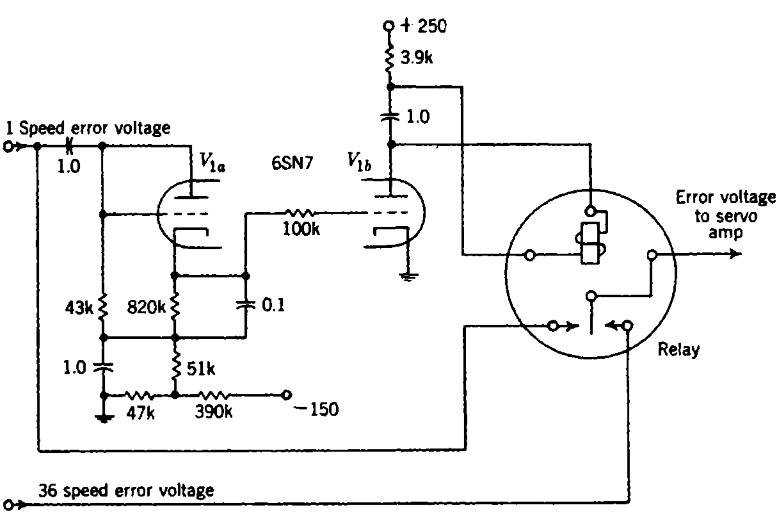


FIG. 12·8.—Amplifier and relay for synchronizing multispeed servo system.

AMPLIFIERS, MODULATORS, AND PHASE DETECTORS[1]

12·9. General Considerations.—The circuits considered here are those which change the voltage that represents the error signal from alternating to direct voltage, or vice versa, or change the voltage scale factor of the error signal. Together with networks and devices for servo stabilization they constitute the remainder of the servo system after the motor, power-control element, and data devices have been selected. The design of the amplifier is related to that of the stabilization networks to such an extent that the amplifier cannot usually be designed entirely independently. One important reason for this interdependence is that gain must usually be sacrificed in stabilization networks; this necessitates higher amplifier gain. Also, the amplifier is often the most logical location for certain stabilization networks.

Loop Gain.—The over-all accuracy requirements determine the loop gain that the servo must have, and these, together with the properties of the control element, data input device, and stabilization networks, determine the gain necessary in the amplifier and associated circuits.

Once a particular value has been assigned to the error that may result from finite loop gain, the necessary gain can be calculated. The usual requirement is that the gain be sufficient to prevent any error larger than the tolerable amount from remaining uncorrected under the given input conditions. If it is necessary to use a d-c error signal, the most critical loop gain factor is usually the drift in level of the first d-c

[1] Sections 12·9 to 12·12 inclusive, are by D. MacRae, Jr.

tube. Error specifications usually require that the drift in the amplifier over any period of time during which it is expected to operate without recalibration must correspond to less than the maximum specified error. Thus, various characteristics of the elements chosen for the servo determine the minimum loop gain that can be used, and this determines the minimum gain that the amplifier and the detector, modulator, and stabilization networks (if any) together must have.

Frequency Response.—The frequency response of the amplifier and phase detector contributes to the over-all frequency response of the loop and influences the stability and transient response of the servo. The most important frequency-response characteristic when a phase detector is used is that due to the filtering or "smoothing" network which follows the phase detector. Choice of phase detector filter characteristics is a compromise between desired attenuation of ripple frequencies and undesired attenuation in the (lower) frequency region that controls the servo stability. The amount of smoothing or filtering necessary is often determined by a vacuum-tube stage to which the filtered output is fed; the alternating component should not drive the tube out of its linear region. This may necessitate a smoothing time constant greater than the period of the carrier. If a low carrier frequency is used—60 cycles, for example—considerable care in design and networks more complicated than single *RC* low-pass filters may be necessary.

Instead of cascading frequency-response corrective networks and amplifiers, one may often combine the networks and amplifiers. Sometimes a desired frequency characteristic may be obtained by use of a feedback amplifier with networks in the feedback circuit. Such a circuit can usually use smaller physical components than the cascaded network and amplifier that it replaces. The amplifier may also be used as a noise filter to reduce the effect of the noise and high-frequency components in the error signal. The theory of corrective and smoothing networks has been discussed in Chaps. 9 to 11.

Factors Determining Type of Circuit Used.—The choice of modulator or phase detector circuits to be used depends largely on the form in which the error signal comes from the data input device and on the type of input required by the control element. The general circuit types may be classified as shown in Table 12·1.

In cases where there is a choice between d-c and a-c amplification, it is usually desirable to use a-c amplifiers because of the additional design freedom resulting from the use of condensers and the consequent ease of setting d-c levels. It is partly for this reason that a-c rather than d-c amplifiers are used in Cases 2 and 3 of Table 12·1. In Case 3, however, a much more important reason for amplifying the alternating voltage before rectifying is the reduction of the effect of d-c drifts. If the phase detector introduces a given d-c drift, this corresponds to a much smaller error if a-c amplification precedes the detector. If the amplifier has gain G, the effect of d-c drift in the phase detector will be less by a factor $1/G$ in the case when a-c amplification is used.

TABLE 12·1.—GENERAL CIRCUIT TYPES

Case No.	Input	Output	Circuit types
1	A-c	A-c	A-c amplifier
2	D-c	A-c	Modulator and a-c amplifier
3	A-c	D-c	A-c amplifier and phase detector
4	D-c	D-c	D-c amplifier; or modulator, a-c amplifier, and phase detector

In the choice between d-c amplifiers and modulation-demodulation systems (Case 4), the latter are preferable, both because of lower drifts and because the difficulties of maintaining levels in a high-gain d-c amplifier are eliminated by the use of a-c amplification. The advantage of d-c amplifiers is economy of parts.

A further consideration in the design of these circuits is the effect of noise or signals other than the desired error signal. This may influence the choice of data input devices in so far as they determine the signal-to-noise ratio when the servo is near equilibrium. If large undesired signals containing harmonics of the error signal frequency are introduced into a phase detector, erroneous output voltages usually result.

12·10. Amplifiers.—A detailed treatment of d-c amplifiers is given in Vol. 18. The design of a-c amplifiers for servos may be done according to well-known methods of audio-amplifier design.[1] Methods of gain control in amplifiers are of special interest to the servo designer. In some cases, such as a resolver servo, the loop gain varies with the input data to the device using the servo; with the aid of some signal also dependent on the input data, automatic gain control may be achieved. An important use of gain controls in servos produced in quantity is to allow near optimum gain adjustment even with large component tolerances or to compensate for varying conditions under which the servos may be used. If tube changes or tube aging produce so large a change in gain that it

[1] See for example: F. E. Terman, *Radio Engineers' Handbook,* 1st ed., McGraw-Hill, New York, 1943; *Radio Engineering,* 2d ed., McGraw-Hill, New York, 1937; H. J. Reich, *Theory and Application of Electron Tubes,* 1st ed., McGraw-Hill, New York, 1939; W. L. Everitt, *Communication Engineering,* 2d ed., McGraw-Hill, New York; and F. L. Smith, *The Radiotron Designer's Handbook,* 3d ed., Amalgamated Wireless Valve Company, Sydney (RCA Mfg. Co.), 1941.

cannot be provided for in the design, an adjustment is usually incorporated so that from time to time the gain may be adjusted for optimum performance. It is sometimes convenient to use a standard servo design with a variety of loads; in this case a gain control may make reasonably satisfactory but not optimum performance possible for a wide range of loads. Gain controls in experimental models are, of course, essential if servos are designed by experimentation, an occasionally justifiable practice.

The usual method of gain control uses a potentiometer to attenuate the signals fed to the grid circuit of one stage of an amplifier. Automatic gain control may be achieved through the use of variable-gain tubes, the gain being electronically controlled by the grid bias voltage. Sometimes controlled limiting is used as a form of gain control; here also the control may be electronic and automatic. The over-all gain of a servo may be varied by operating on the motor or power control element. While the principal gain control methods have been listed, others will suggest themselves in special applications.

A "balance control" is often inserted in the servo loop to ensure that for zero error signal the output signal to the motor is zero. A common method is to insert a potentiometer between the cathodes of a differential amplifier or phase detector, the cathode resistor being connected to the movable contact of the potentiometer. An adjustment may then be made in the servo loop. Where feasible, it is preferable to omit the balance control, using instead a loop gain high enough so that the expected unbalance requires a compensating error of a magnitude that is small compared with the allowable error.

12·11. Modulators. The Use of Modulators in Servos.—A modulating device used in a servo must provide an alternating output voltage that indicates the sense as well as the magnitude of the input signals; if the polarity of the input changes, corresponding to a change in the sign of the error, this must be indicated in the output. The most common method of indicating change in input polarity is an inversion of the output waveform. This symmetry condition somewhat restricts the class of modulator to be considered in connection with servos.[1] The conditions will also be imposed that output amplitude be roughly proportional to input amplitude, at least in the vicinity of zero error signal, and that for zero d-c input, there be zero a-c output. While the input signal to a modulator device may be a single-ended d-c error signal, it is more commonly two d-c signals whose difference represents the error signal. The class of circuits most frequently used as modulators uses elements that act as switches: diodes, crystal rectifiers, triodes used as clamp tubes, mechanical

[1] For a more general discussion of modulator circuits, see Vol. 19 of this series, and Secs. 3·10 and 15·4 of this volume.

contactors, etc. The general property of such circuits is that during one portion of the cycle of the carrier certain switches are closed and others are open whereas during the remainder of the cycle the switch positions are reversed. In the circuits using two-terminal devices (e.g., diodes, crystal rectifiers) as switches, both the modulating signal and the carrier in general influence the time at which the switches may be considered to open and close; but if the carrier amplitude is large, as is usually the case, the opening and closing times may be considered functions of the carrier waveform alone.

Analysis of Switch-type Modulator and Phase Detector.—It may be shown easily that any modulator of this type may also be used as a phase detector. Suppose the circuit consists of double-throw switches that have one position during one part of the carrier cycle (the part having duration t_1) and another position during the remainder (of duration t_2) of the cycle. Assume further that the switches are "ideal" and that during either interval t_1 or t_2 the circuit may be considered a linear resistive network. A modulator of this type can then be described as a device that produces an output e_2 which varies as a function of the input e_1 in the following manner:

$$
\begin{aligned}
e_2 &= A_1e_1 \qquad \text{during the interval } t_1,\\
e_2 &= A_2e_1 \qquad \text{during the interval } t_2.
\end{aligned}
$$

If e_1 is a slowly varying voltage over the interval $t_1 + t_2$, the device then acts as a modulator; the output will be an alternating voltage of rectangular waveform whose peak-to-peak amplitude is $A_1e_1 - A_2e_1$. The output therefore has the desired property of inverting when e_1 changes polarity. A device described by these equations can also be used as a phase detector if the input e_1 is a periodic waveform of period $t_1 + t_2$. If the waveform e_1 varies in amplitude but maintains its shape, it can be described by the expression

$$e_1 = Cf(t), \tag{1}$$

and C is a real quantity that varies slowly over the interval $t_1 + t_2$ and can be positive or negative. If the average values of $f(t)$ during the intervals t_1 and t_2 are $\bar{f}_1$ and $\bar{f}_2$ respectively, then the average output voltage of the circuit is

$$e_2 = A_1C\bar{f}_1 \tag{2}$$

over the interval t_1, and

$$e_2 = A_2C\bar{f}_2 \tag{3}$$

over the interval t_2. Consequently the average output voltage over the entire cycle will be

$$\bar{e}_2 = C\left[\frac{A_1\bar{f}_1t_1 + A_2\bar{f}_2t_2}{t_1 + t_2}\right], \tag{4}$$

and the average output will vary with C if the quantity in brackets does not vanish.

Practical Circuits.—An excellent example of this type of circuit is the mechanical switch,[1] which may be used either as a modulator or as a phase-sensitive demodulator. If the switch short-circuits the signal to ground during the interval t_2 to a good approximation, one may write for conventional modulators

$$t_1 = t_2,$$
$$A_1 = 1,$$
$$A_2 = 0. \qquad (5)$$

The output waveform is a rectangular wave with peak-to-peak amplitude equal to the input e_1. If the same circuit is used as a demodulator,

$$\bar{e}_2 = C\bar{f}_2 \qquad (6)$$

The averaging process to separate $\tilde{e}_2$ from higher-frequency components may be done with a simple RC low-pass filter or other type filter.

Four typical circuits using switches for modulation are shown in Fig. 12·9. The output waveforms of the shunt and series circuits

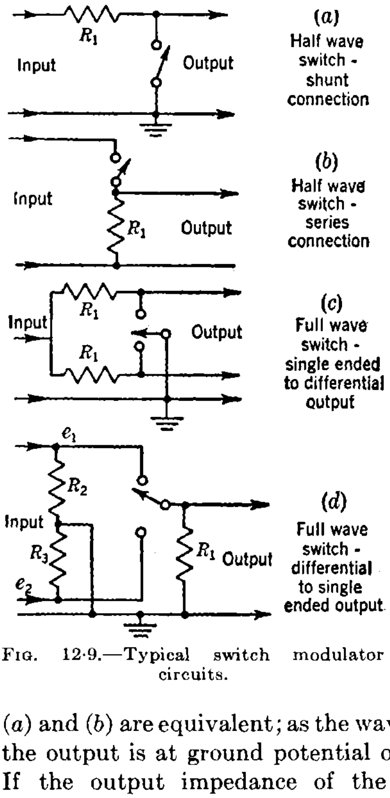


FIG. 12·9.—Typical switch modulator circuits.

(*a*) and (*b*) are equivalent; as the waveform photograph (Fig. 12·10) shows, the output is at ground potential on alternate half cycles of the carrier. If the output impedance of the source is high, circuit (*a*) is preferable; if it is low, circuit (*b*) can be used. C rcuit type (*c*) produces a push-pull modulated output. Circuit (*d*) converts a push-pull or differential signal to a single-ended output; it is particularly useful, since the comparison of two d-c voltages is usually necessary and most of the a-c amplifiers used for the modulated carrier are single-ended.

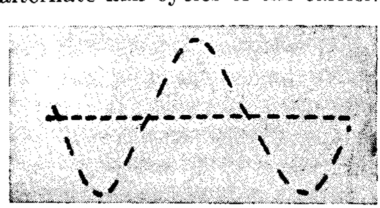
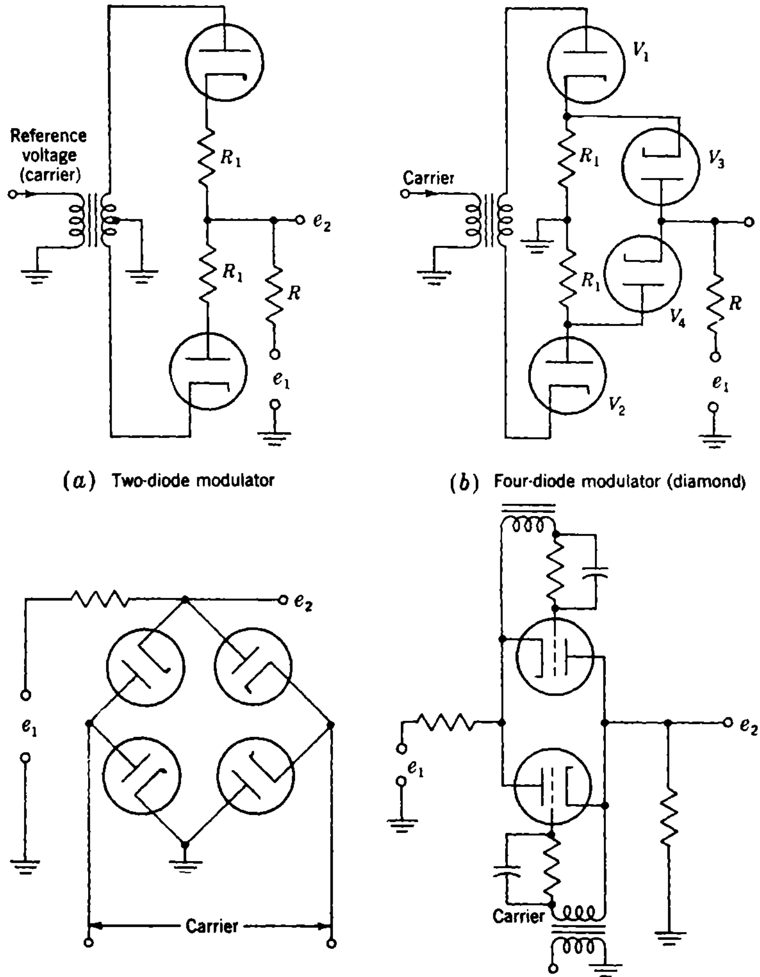


FIG. 12·10.—Waveform of switch modulators of Fig. 12·9 *a* and *b*.

[1] See Chap. 16 for more details on switch modulators.

Several other modulator circuits of the switch type are shown in Fig. 12·11. In the first three, diodes or other two-terminal unilateral



(*a*) Two-diode modulator
(*b*) Four-diode modulator (diamond)
(*c*) Bridge-type modulator
(*d*) Double-triode clamp as modulator

FIG. 12·11.—Switch-type modulators that may also be used as phase detectors.

impedances (copper oxide, selenium, or germanium rectifiers, for example) may be used. In the fourth the two triodes together constitute the electronic equivalent of a switch.[1]

[1] Other switch-type modulator circuits are shown in F. E. Terman, *Radio Engineers' Handbook*, McGraw-Hill, New York, 1943, p. 553. See also Vol. 19 of this series.

The two-diode modulator of Fig. 12·11*a* effectively clamps the output to ground every other half cycle. When the voltage at the upper end of the transformer is positive, both diodes conduct, the current through them being limited by the resistances R_1. If $R_1 \ll R$, the output e_2 will be nearly zero during this interval. When the carrier has the opposite polarity, the diodes are effectively open circuits and $e_2 = e_1$. The principal errors in the operation of this circuit result from the fact that the diode characteristics are not ideal and identical. The departure of diodes from the ideal characteristic is treated in detail in Vol. 19. Asymmetry may cause sharp positive or negative "spikes" to appear in the output waveform and obscure the null that should be produced when the d-c input is zero. The reduction of heater voltage produces improvement in some cases at the expense of shortened tube life. The four-diode modulator ("diamond") shown in Fig. 12·11*b* shows less dependence on diode characteristics. During the half cycle when the upper end of the transformer secondary is positive, diodes V_1 and V_2 conduct while V_3 and V_4 are cut off, and $e_2 \approx e_1$. On the opposite half cycle, V_1 and V_2 are cut off and V_3 and V_4 connect the output to ground through resistors R_1. This has the advantage that the null point of the modulator is determined by the diode V_3 and V_4 at only one part of their characteristics; in Circuit (*a*) symmetry was required over a considerable range of the conducting regions.

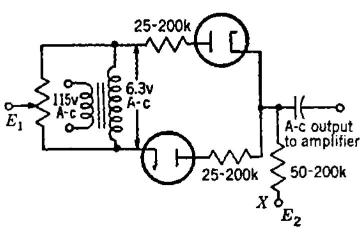


FIG. 12·12.—Modulator comparing two d-c voltages.

The circuit of Fig. 12·11*c* is the same as Circuit (*a*) of Fig. 12·9, except that a four-diode bridge is used in place of the mechanical switch.

Most of the modulator circuits of Fig. 12·11 may also be used when the difference of two d-c voltages is the modulating signal. In this case one of the inputs is at the same input terminal as in Fig. 12·11, and the other is at the ground terminal. Circuit (*a*), for example, might assume the form shown in Fig. 12·12. The additional input has now been added effectively at the center tap of the transformer. In case the transformer winding is not center-tapped with sufficient symmetry, or if a transformer has no center tap, a potentiometer in parallel with the transformer secondary may be used as shown to produce the same effect. When the transformer secondary is floating as in this case, it is desirable to use a well-shielded transformer, for commercial transformers frequently have high capacitive coupling between windings. As a result of this coupling, an undesired alternating voltage of the carrier frequency may be superimposed on the d-c input. This effect is particularly noticeable at 400 cycles and higher carrier frequencies; if generators are used, their high harmonic content accentuates the effect. In order to minimize this effect, the designer should keep input impedances to ground low or introduce an *RC* low-pass filter in one or both d-c inputs.

Modulators of other varieties than the switch type may also be used. Those employing special characteristics such as double input vacuum tubes and those employing signal-controlled amplitude selection[1] do not in their simplest forms provide zero a-c output for zero input. This can be remedied, however, by the construction of a balanced circuit in which one output is subtracted from another produced by a similar element with the signal or carrier inverted. One example of such a circuit is shown in Fig. 12·13.

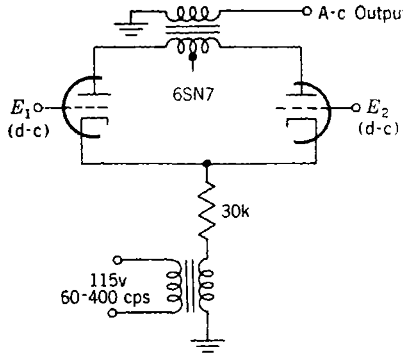


FIG. 12·13.—Nonswitch-type modulator.

12·12. Phase Detectors.—To convert a modulated alternating voltage, measuring an error signal, to a direct voltage requires a demodulator that responds to the sense of the modulated wave.[2] Such circuits are commonly called "phase detectors," although in servo applications it usually is the variation of amplitude from positive to negative values rather than the variation of relative phase angle that is of importance. Most of the circuits shown will detect either phase or amplitude variations, however. Phase detectors for servos must have relatively stable zero points and must remain "saturated" for very large input signals corresponding to large servo errors. The circuits mentioned above, using switches that operate at times determined by the carrier or reference signal and providing no amplification in themselves, belong to the important and useful controlled-switch phase detector class. Other types of phase detectors of importance are those in which switching does not occur at fixed times and those which amplify as well as provide switching.

A phase detector in which switching does not occur at fixed times is shown in Fig. 12·14. It operates on the principle of adding the error signal to and subtracting it from the reference voltage. The two resulting alternating voltages are rectified; the two d-c voltage outputs are equal only when the component of error signal in phase with the reference voltage is zero. A similar circuit is used in some f-m discriminators.

A number of phase detector circuits that use triodes and provide

[1] These modulator circuits are treated extensively in Vol. 19.
[2] See Vol. 19.

amplification are shown in Fig. 12·15. In each of these circuits, two triodes are balanced against each other to give symmetrical operation. It is necessary to use either a push-pull signal or a push-pull reference voltage (carrier), for if the signal is zero, the average currents through the triodes are the same, and cancel by subtraction; but as the error signal varies, the subtracted currents will then vary in opposite directions. In symmetrical circuits of this kind, differences in tube characteristics may lead to asymmetry in the output at zero signal. This effect is less important in phase detectors than in modulators, however, for phase detectors

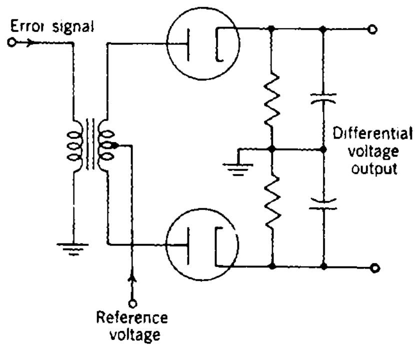


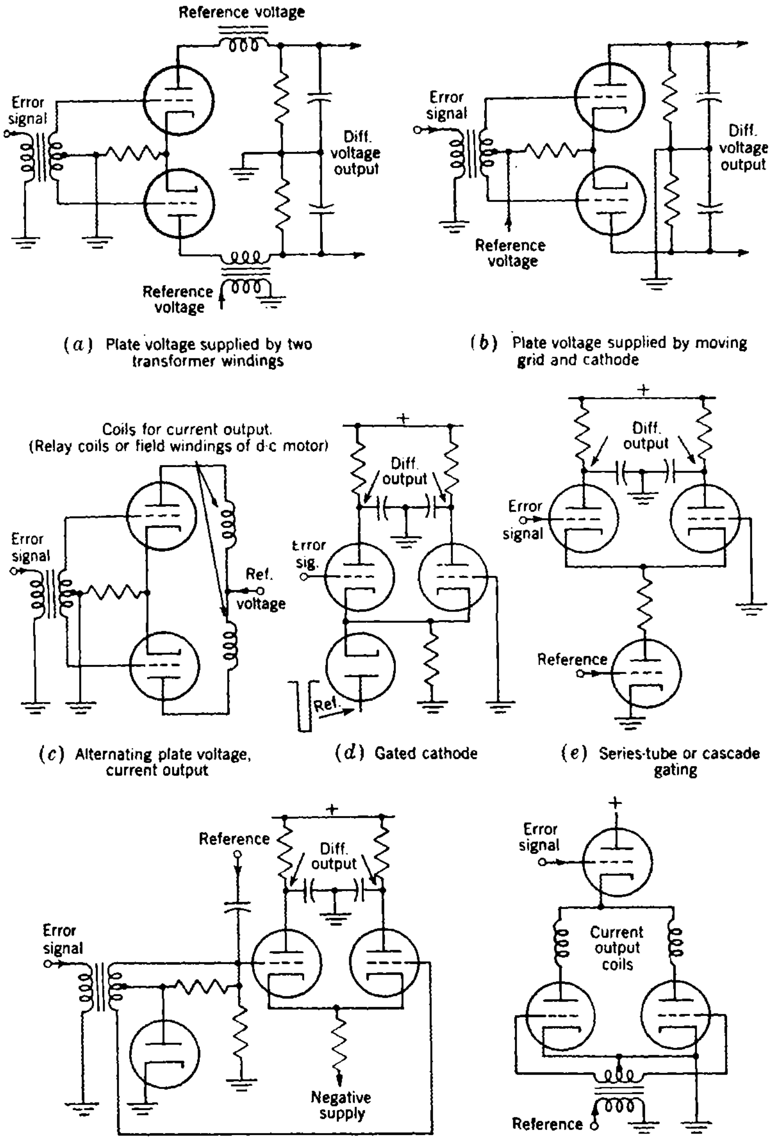
FIG. 12·14.—Phase detector using addition of voltages.

are usually operated at voltage levels that are high relative to the voltage asymmetry which may result from tube differences.

Circuit (*a*) uses a push-pull signal at the grids of two triodes and provides the two plate voltages by means of floating transformer windings. Circuit (*b*) is similar but provides the plate voltages by changing the voltages of the grids and cathodes relative to the plates. In Circuit (*b*) an undesired signal may result from capacitive coupling between the secondary of the transformer and ground.

Circuit (*c*) resembles Circuit (*b*) with the exception that the output current is used directly to control relays or a motor. If the windings in which the current is used may float with respect to ground, the reference voltage may be applied as shown without a transformer. Either (*a*) or (*c*) may be used with push-pull plate voltages and a single-ended signal which is fed to both grids; in this case a transformer is necessary to supply the plate voltage.

In Circuit (*d*) the switching action, or time selection, is made more definite by the use of a rectangular reference voltage. When the reference waveform is positive, the cathodes of the differential amplifier are raised and the tubes are cut off; during the negative part of the reference waveform, the diode does not conduct and the differential amplifier func-



(*a*) Plate voltage supplied by two transformer windings

(*b*) Plate voltage supplied by moving grid and cathode

(*c*) Alternating plate voltage, current output

(*d*) Gated cathode

(*e*) Series-tube or cascade gating

(*f*) Gated grid

(*g*) Alternate cathode gating with current output

Fig. 12·15.—Phase detectors with amplification.

tions as an amplitude discriminating circuit. The common cathode resistor effectively provides nearly push-pull grid-cathode voltages. This phase detector is used in the a-c integrating servo described in Sec. 4·10.

Circuit (*e*) uses a series tube, which is effectively a switch, in series with a differential amplifier. When the reference waveform is positive, the differential amplifier functions normally; when it is negative the switch tube and consequently the amplifier tubes are cut off.

Circuit (*f*) is similar to (*d*), since it is a differential amplifier which operates normally during part of the carrier cycle. In this case an approximation to a rectangular reference waveform may be generated from a sine wave by means of the resistance-capacitance-diode circuit.

Circuit (*g*) is similar to a cathode follower with alternate paths for the cathode current; it provides a differential current with low output impedance. This circuit has the advantage of maintaining its zero fairly well with respect to component variation.

In the circuits using sinusoidal reference voltages and error signals, difficulties may arise from phase shifts of the error signal. A phase shift may cause a slight overlap between the reference voltage and the error signal in the "off" tube. During the interval when the error signal and reference voltage coincide, a pulse of current will flow in this tube; and if the error signal is large, the tube may be turned "full on" (driven to zero grid bias). If a smoothing filter is used which gives peak detection, the output of the "off" tube may be nearly the same as that of the "on" tube. This can result in sluggish servo action at large values of error.

The amount of filtering necessary in the output circuit of a phase detector depends largely on the nature of the circuit to which the output voltage is fed.

MOTORS AND OTHER POWER DEVICES

12·13. Motors.[1]—The design factors listed in Chap. 19 apply to the choice and design of motors. In connection with this list, attention is called to the following factors: type of power available, commutator or slip ring electrical noise, brush wear, output power, speed range, torque, inertia, and smoothness.

It must be emphasized that most simple servo design problems can be solved with any one of a number of motors. Usually the choice of one motor in preference to others is intimately related to the choice of the rest of the circuit, and particularly the power-control portion of the circuit.

Commutator electrical noise is frequently a serious source of interference with other equipments. A-c devices using slip rings or devices

[1] Section 12·13 is by I. A. Greenwood, Jr.

using no rotating contacts at all are considerably superior to d-c motors on this count. Commutator noise can nearly always be eliminated—it is merely a painstaking job to do this. Usually electrical noise gets out of the motor by radiation and by conduction on the power leads; both shielding and filters may be required to eliminate such motor noise effectively.

The type of power available will affect the selection of a motor. It is necessary to know whether the power source is alternating or direct current and to know the voltage, the frequency, the load that may be drawn, and sometimes the waveform.

D-c motors should be considered from the standpoint of brush wear, particularly in the case of aircraft equipment which must operate at high altitudes. Care should be taken to see that brush life is adequate for the applications and that carbon dust from the brushes does not provide undesirable leakage paths. Frequently a-c motors are chosen in preference to d-c motors just because maintenance is easier.

Power and torque-to-inertia ratio or torque-to-square-root-of-inertia are very important characteristics of motors with regard to their use in servos. Power determines the load that can be put on a servo, while the torque-to-inertia or torque-to-square-root-of-inertia ratio affects the upper limit of the speed of response (acceleration) of a servo containing the motor, along with power and smoothness. It was shown in Sec. 11·12 that where gear ratios can be changed, maximum acceleration at the load is produced by the motor having the highest torque-to-square-root-of-inertia ratio; where gear ratios are predetermined and large, the maximum acceleration is produced by a motor having largest torque-to-inertia ratio. For fastest response, specially designed squirrel-cage induction motors, drag-cup induction motors, or magnetic-clutch devices should be considered. The smoothness of a motor greatly affects its use in a servomechanism, since it is a factor in determining how small a gear reduction can be used. Smoothness is improved by armature skewing, thus reducing slot lock; many armature poles and commutator bars; reduction of friction; and reduction of backlash.

It is a fundamental feature of feedback technique that over-all characteristics of a circuit with feedback are less dependent on the characteristics of the μ portion of the loop and more dependent on the β portion of the loop. In a servo system, the β portion, consisting usually of data input and output devices and associated circuits, may be made very accurate, linear, and stable compared with the elements of the μ portion of the loop. Because high-gain feedback loops are involved in most servos, and because the motor is part of the μ portion of the feedback loop, motor and motor-control characteristics are not very critical. There is another factor in the operation of many servos that tends to make the motor characteristics less critical. In these servos control of power to the motor is accomplished by controlling the fraction of the time that full power is applied to the motor rather than controlling the amount of power applied steadily to the motor. This means that as compared with nonservo operation of motors at low speeds the effects of irregular friction, slot-lock, etc., cause less variation in operation, since they are acting against full motor torque. This type of power control is available in circuits such as relay control, thyratron control, and other types. In circuits where genuine continuous control is used, the slight changes in motion caused by the load and friction irregularities listed also cause short pulses of power to be applied to the motor, thereby making full torques available at low speed. By way of comparison, when the ordinary motor is run at low speed without being part of a closed-cycle control system, a slight increase in torque load may stall the motor, while a slight decrease in torque load will cause it to speed up with the result that operation at low speeds may be very unstable and jerky and completely impossible at the very low speeds at which servos may be smoothly operated. This technique of applying short pulses of power to a motor has a disadvantage that must be considered: The a-c component of the resulting drive waveform produces heat but no average torque in those motors where the alternating current does not also act on the field. Dissipation of this a-c energy must be provided for.

D-c Motors.—D-c motors in general have the advantages of being lighter for the same power output and of having higher starting and reversing torques than a-c motors.

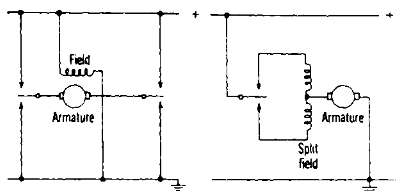
Series motors have high starting torque and poor speed regulation with torque. This high negative slope of the torque-speed characteristic (the latter is approximately linear) is equivalent to a high viscous damping torque and contributes materially to servo stability[1] but adds also to velocity errors. Higher torques on reversal can be obtained with series motors than with other types of motor. In general, a straight-series motor has the disadvantage of being a unidirectional device unless some type of switching is used that reverses either the armature or field connections but not both. Series motors with such change of field connections made by a relay have been used in some cases to achieve bidirectional control, an example being a radar trainer integrator servo. This type of relay-switched direction control is suitable for velocity servos where the direction is not changed frequently and where operation very close to zero speed is not required but would not be very suitable for a position control.

Split-series motors are characterized by high starting torque and have the advantage of bidirectional control with a small number of control

[1] See Sec. 13·2.

elements. In Fig. 12·16 are shown two typical relay control circuits, one for a separately controlled armature motor and one for a split-series motor. It is seen that only half as many relay contacts are required for the split-series motor. Because the split-series motor can utilize only one-half of the field winding space at any one instant, its design is generally not so compact as a straight-series motor of the same rating. Comparing motors wound on the same frame, a split-series motor will have a lower field flux and a lower torque rating than a straight-series motor.

Shunt motors and armature-controlled motors are characterized by a smaller change in speed with load than series motors. Starting torques are moderate. Shunt motors used in servos are generally operated as armature-controlled motors, with the field windings continuously connected to the power source. When so connected, these motors may be reversed by reversing the armature current. If operated as straight shunt motors, reversal of the direction of rotation must be done by reversing armature or field leads but not both, as with series motors. By taking advantage of the intermittent duty ratios of most servos, the torque limitations of shunt motors may be overcome to some extent by the application of momentary overloads.



Relay bidirectional control separately controlled armature motor. Two relay poles are required.

Relay bidirectional control of split-series motor. Only one relay pole is required.

FIG. 12·16.—Reduction of relay pole requirements by use of split-series motor.

Permanent-magnet motors are simple and light but are subject to the limitation that the field may be demagnetized if the motor is badly overloaded. Specifically, demagnetization of the field by the armature is likely to occur if the demagnetizing force due to armature current exceeds the value for which the magnet is stabilized. Frequently, rotation of the poles with respect to the brushes rather than demagnetization is observed. This limitation, once recognized, is not serious, for it is relatively easy to guard against the possibility of such overloads when designing driving circuits.

In circuit applications, permanent-magnet motors are characterized by economy of drive, in that no power is used in the field. By reversing the direction of armature current the direction of rotation of the motor is easily reversed. High torques on reversal are available.

Field-controlled motors have the advantage of high-impedance drive, which makes them particularly suitable for control with small vacuum tubes. Ordinarily, when controlling a motor by controlling its field, the field current must be increased for a decrease of speed. This characteristic of motors has in general made field control of d-c motors unpopular. However, if the armature current is kept reasonably constant rather than allowed to vary over wide limits as the armature back emf changes due to changes in field current, then it is possible to achieve a satisfactory motor control in which there is speed or torque increase with increase

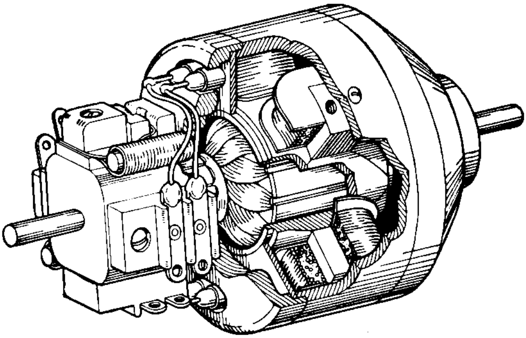


FIG. 12·17.—Lightweight field-controlled d-c motor, Holtzer Cabot RBD 0808.

of field current. Three examples of such motors[1] are the British Velodynes, the MIT Servomechanisms Laboratory 100-watt instrument servo motor, and the RBD 0808, a lightweight aircraft instrument motor designed for the Radiation Laboratory by the Holtzer-Cabot Company (Fig. 12·17). All these motors may be controlled by vacuum tubes whose power ratings are small compared with the power of the motors; i.e., the motors themselves constitute power-amplifying devices. The inductance of the field of such motors is usually very high. In this type motor, field hysteresis effects must be kept to a minimum; otherwise serious instability may result comparable to that arising from excessive backlash.

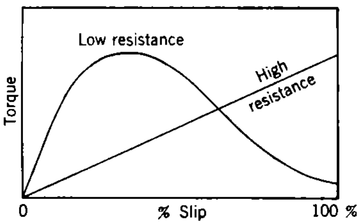


FIG. 12·18.—Effect of armature resistance on characteristics of induction motor.

Alternating-current Motors.—Alternating-current motors are extensively used in servo applications because of their simplicity, reliability, absence of commutator sparking, rapid response, and economy.

Two-phase induction motor characteristics depend on the resistance of the conducting material of the armature. If this resistance is low, the torque vs. slip characteristic is as shown in Fig. 12·18 by the curve marked "low resistance." The result of this type of characteristic is low starting torque but higher torque near top speed. This is true in

[1] See Sec. 12·19 below.

most induction motors. By increasing the armature resistance, the starting torque may be increased at the expense of torque at high speeds. This may be done by the use of such materials as zinc for conducting bars. Alternating-current motors for servo applications are generally designed so that maximum torque occurs with the motor stalled, although maximum power is usually near top speed. As with series motors, this char-

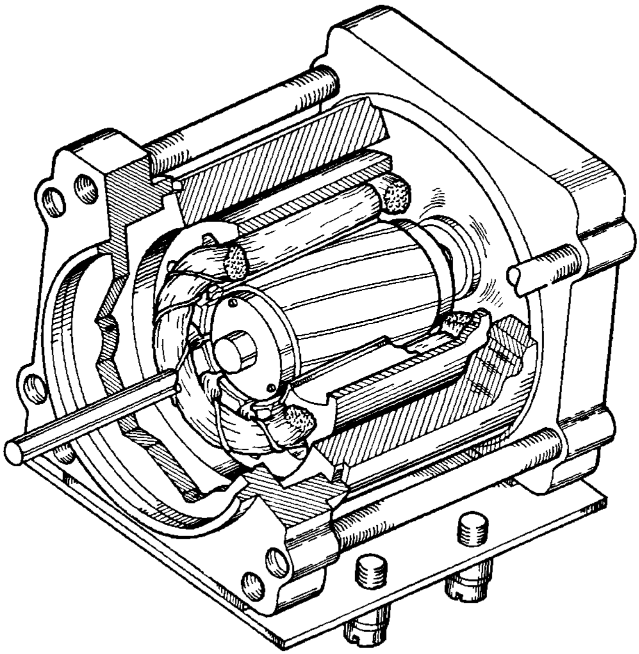


FIG. 12·19.—Cutaway view of a typical Diehl low-inertia squirrel-cage induction motor.

acteristic of decreasing torque with increasing speed contributes to servo stability.[1] As in the case of shunt motors torque may be increased merely by momentary overload up to the point where saturation of the iron becomes a limiting factor.

The rapid response of some induction motor servos is due to their high torque/inertia and torque/square-root-of-inertia ratios. Inertia can be kept low because of the simplicity of the armature.

In both the Diehl low-inertia squirrel-cage induction motors (Figs. 12·19 and 12·20) and the drag-cup type induction motors, torque-inertia ratios may be very high, accelerations of greater than 6000 radians per second per second being possible with both motors. While acceleration

[1] See Sec. 13·2.

figures for the drag-cup type motor are somewhat higher than for the squirrel-cage motor, this type of motor is characterized by low torque for its size and weight.

It is more efficient for the two windings of an induction motor to be of different impedance. The low-impedance winding is connected directly to the a-c power source, while the high-impedance winding is connected to the control circuit. Direction of rotation is changed by change of phase

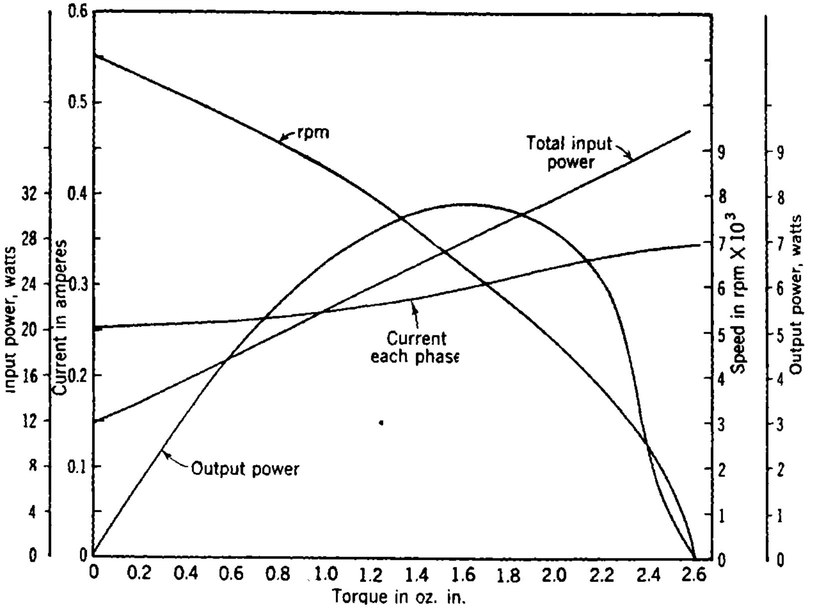


FIG. 12·20.—Characteristics of Diehl FPE-25-9 low-inertia squirrel-cage induction motor.

in the control winding. A phase shift of 90° must be introduced in the current flowing in either the line phase or the control phase. In practice both methods are used. Some induction motors, such as the Bendix-Pioneer CK-5, have eddy-current damping devices built into the motor. This type of damping allows a-c servos to be built in which the error signal need not be converted to direct current. Stabilizing circuits for a-c signals are not very satisfactory unless the frequency of the signal is kept relatively constant; in the applications where this motor is most frequently desired, the required constancy of frequency would be impractical. For applications where highest accuracy is required, damping of this circuit is usually supplemented by feedback from an a-c tachometer.

Synchronous motors are used mainly as a means for data transmission and as synchronous drives for such mechanisms as timers. They are seldom if ever used for servo-motor drives.

Universal motors are series motors that may be operated on either alternating or direct current. In order to achieve bidirectional control either the armature or field windings must be reversed but not both.

Single-phase a-c motors require special tricks in order to achieve reversal of direction of rotation. Fixed pole shading is generally used to cause the motor to start in the desired direction. Bidirectional control may be achieved by using two sets of shading coils on the pole, each tending to cause rotation in an opposite direction, and remotely completing the circuit of the appropriate coil leaving the other coil open-circuited. This makes an extremely cheap and easily reversed motor.

The split-phase motor is similar to the two-phase induction motor, in that the current in one winding is shifted in phase by a capacitor, usually connected in shunt with the high-impedance winding. In this case the impedance of the phase-shifted winding is larger than the directly fed winding, so that a small capacitance may be used. Unless connections to either winding, but not both, are reversed a split-phase motor is a unidirectional device.

Repulsion motors resemble series motors in their characteristics, in that they have high stall torques and high speeds at light loads with full power applied. Repulsion motors may be controlled for servo applications by control of current in the path between the shorting brushes. By the use of two sets of brushes for each pair of stator poles, bidirectional rotation can be achieved. This type of control is used in the MIT differential analyzer.[1]

12·14. Magnetic Clutches.[2]—The term magnetic clutch is applied to a family of devices used to control the coupling of torque from an input shaft to an output shaft by means of electrical signals. The torque coupling may be achieved by contact between friction surfaces or by interaction of a magnetic field and the field associated with eddy currents produced by the motion of the first-mentioned field past a conducting surface. Since a magnetic clutch is required only to transmit torque rather than to generate torque, it may have power and torque outputs that are large for the size of the device and the power needed to control it. Since the input shaft to a magnetic clutch may run at a constant, relatively high speed, high momentary torques may be made available to the clutch by the simple expedient of using a large inertia on the input shaft. The result of all this is that a small motor turning at high speed a large flywheel coupled by a magnetic clutch to a load may be able to impart very large acceleration to the load—much larger than is possible with just an electric motor—and an equal or somewhat larger accelera-

[1] V. Bush and S. Caldwell, "A New Type of Differential Analyzer," *Jour. Franklin Inst.*, **240**, October 1945.

[2] Sections 12·14 to 12·16, inclusive, are by J. R. Rogers and I. A. Greenwood, Jr.

tion as compared with a hydraulic motor. For applications where very high accelerations or very large torque overloads are involved, magnetic clutches are therefore of great interest.

There is another feature of magnetic clutches that partially justifies their use in some other applications, namely, the effective power amplification that may be achieved. The practical statement of this property is that magnetic clutches capable of supplying instrument servo power may be controlled from small thyratrons or receiving-type vacuum tubes.

Magnetic clutches are used in two distinct ways: as electrically controlled connections between mechanical shafts and as power devices in servomechanisms. The latter usage will be the chief interest of this section. The former usage should not be forgotten, however, for it has many applications in electronic and electromechanical devices, a typical example being its use in the MIT differential analyzer.[1]

Friction vs. Eddy-current Types.—Two distinct types of magnetic clutches are recognized, those which transmit torque by the physical contact of frictional surfaces and those which transmit torque by the action of magnetic flux produced either by two sets of coils or by one set of coils and eddy current induced in a conducting surface by the rotation of this set of coils. The eddy-current variety is more common than the two-coil variety. It is of interest to compare the frictional surface and eddy-current types of magnetic clutches on the basis of wear, heat dissipation, ability to give proportional control, and power amplification. The problem of wear is a serious consideration in the design of clutches operating by contact of frictional surfaces. Allowance for excessive wear may require an unduly large air gap and thus complicate the design of the magnetic structure. For instrument-size magnetic clutches, cork on steel has been found to be a reasonably satisfactory combination of surfaces and is probably used on more equipments than any other combination. For larger clutches there are available a variety of materials which have been developed especially as clutch faces. From the standpoint of heat dissipation there is little basis for choice between the two types of clutches. When both types are operated with the same slip, the heat dissipated is principally that due to the conversion into heat of the power represented by the product of torque and slip and the I^2R loss in the control coil. Under the same conditions of torque transmitted and slippage, the two types will, of course, have equal torque-slip power dissipation. The difference in I^2R losses is small but will be in the favor of the frictional surface type, for this type requires less power to operate than does the eddy-current type. In the eddy-current type clutch, some additional cooling may be achieved as compared with

[1] Bush and Caldwell, *op. cit.*

the friction surface type by the circulation of air through the air gap, where it may directly contact the region of origin of much of the heat. Forced circulation is seldom worth its cost in magnetic clutches designed for instrument servo applications. From the standpoint of smoothness of control, the eddy-current type magnetic clutch has a considerable advantage over the frictional surface type. Factors controlling the lack of smoothness of frictional surface type clutches are the difference between static and coulomb friction, the amount of "lumpy" or irregular friction, and the tendency of the frictional surfaces to vibrate or bounce. These effects are not present, of course, in the eddy-current type clutch. In spite of the effects just mentioned, a frictional type clutch may be made smooth enough for most applications. In this type clutch, decrease in the size of the air gap as exciting power increases results in a torque characteristic that increases more nearly as the square of the exciting current than as the first power. In the eddy-current type device, the magnetic air gap stays constant; but for a given value of slip, both the exciting flux and the flux produced by the eddy currents increase as the exciting current increases, with the result that the torque output is roughly proportional to the square of the exciting current. Thus, there is little choice between the two types from the standpoint of proportionality of control characteristics; both are sufficiently close to linearity in their characteristics to be entirely suitable in this respect for the less demanding servo applications.

Proportional Control.—One additional method of achieving proportional control with clutches (principally of the frictional surface type) which may act in an essentially "on-off" manner is through the use of "buzz" circuits. "Buzz" control of relays for controlling servo motors is discussed in Sec. 12·19. The principles discussed in this section apply directly to the present section.

Moving-coil vs. Stationary-coil Types.—Magnetic clutches are made in both the moving-coil and the stationary-coil type. In the moving-coil type, the coil actuating the moving member rotates, and leads are connected to it by means of slip rings. In the stationary-coil type, an extra air gap is introduced, so that both the driving member and the driven member may rotate with respect to fixed coils. Examples of both types are shown below.

Velocity, Torque, and Acceleration Considerations.—In Fig. 12·21 is shown a gearing and inertia schematic for a typical magnetic clutch installation. For a given application, the gear reduction n_1n_2 is determined by the ratio of the speed of rotation of the motor and the desired speed of rotation of the output shaft. For a given torque applied to the load under these conditions, it is of interest to note that the relative values of n_1 and n_2 do not affect the power dissipated in the clutch.

It will be recalled that the power dissipated in the clutch is nearly equal to the I^2R loss in the windings and the torque-slip product. If in Fig. 12·21 n_2 is increased to n_2' and n_1 is decreased to n_1', where $n_1'n_2'$ equals n_1n_2, the torque that the clutch must transmit is reduced by the factor n_2/n_2', while the slip of the clutch is increased by the ratio n_1/n_1'. Thus the product of torque and slip remains constant, and other criteria may be used in choosing the relative values of n_1n_2, given their product. It is, of course, necessary that the torque T transmitted by the magnetic clutch when multiplied by the last gear ratio n_2 be great enough to furnish the required output torque.

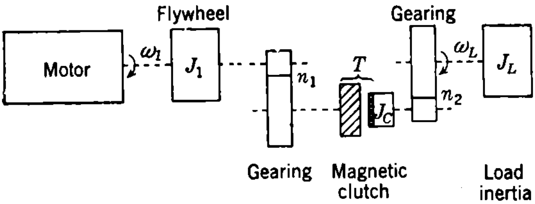


FIG. 12·21.—Mechanical schematic of magnetic clutch drive.

The choice of the gear ratio n_2 to produce the maximum acceleration of the load inertia J_L will next be considered. This subject is discussed in Sec. 11·12. Using the methods of that section, it is easily shown that a maximum angular acceleration of $T/2\sqrt{J_cJ_L}$ will result when the gear ratio n_2 is chosen such as to make equal the inertia of the load-side member of the clutch J_c, as seen at the load, and the load inertia J_L. It will be recalled that inertia seen through gearing varies as the square of the gear ratio. Thus, for maximum acceleration at the load, $n_2^2J_c = J_L$. Consideration of the equation for maximum acceleration leads to a simple criterion which may be used as a figure of merit in comparing clutches. This figure of merit is the transmitted torque divided by the square root of the inertia of the output member.

On the basis of the relations just given, a simple procedure can now be formulated for using the various constants involved. Let it be assumed that the load inertia J_L and the maximum required load velocity ω_L are given and that a servo response is to be as fast as possible; i.e., the load acceleration is to be made as large as possible. A design may be started by choosing or designing a magnetic clutch with a value of $T/\sqrt{J_c}$ as large as is possible or reasonable. The gear ratio n_2 is then chosen to match the clutch inertia J_c and the load inertia J_1, by use of the relation $n_2^2 = J_L/J_c$. A motor is next chosen capable of supplying the necessary average power. In order that a small flywheel may be used with the motor, it is preferable to choose a motor having a high speed and low torque rather than one having the same power but with low speed and high torque. The gear ratio n_1 is then chosen so that $\omega_1/n_1n_2 = \omega_L$ for the condition of zero slip.

The flywheel inertia J_1 which also includes the inertia of the motor armature is next chosen such that $J_1n_1^2$ is large compared with $2J_c$. The design procedure given is, of course, only a part of the complete design procedure, which is discussed in more detail in Chap. 8. Further methods and equations for the design of magnetic clutches are given by Andrews and Shanely[1] and in Vol. 17 of this series.

Typical Designs and Applications.—Magnetic clutches have been used extensively in servomechanisms used to control aircraft. A familiar example of this type of usage is the Minneapolis-Honeywell C-1 automatic flight control equipment.[2] This equipment uses a total of four frictional-type magnetic clutches, in combination with a differential. A motor running at constant speed is coupled by a magnetic clutch to one of two input shafts of the differential, the other input shaft being locked by means of a magnetic clutch. By this means, bidirectional rotation of the output is achieved. When either clutches or brakes are not energized, the output may be turned through the manual aircraft controls without additional load, even though the servo drive motor is running. The frictional surfaces used in these clutches are cork on polished steel.

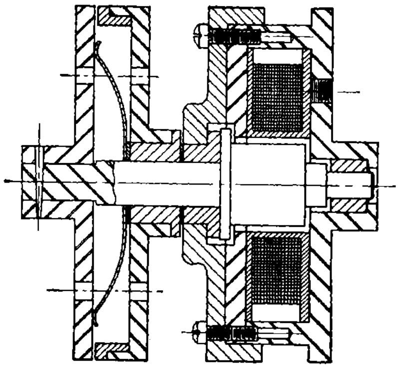


FIG. 12·22.—Fixed-coil friction-type magnetic clutch.

Figures 12·22 and 12·23 show designs for small friction-type magnetic clutches.

The clutch shown in Fig. 12·22 is designed to have a low inertia and has a maximum torque of 15 oz-in. Both clutches are controlled by 0 to 27 volts applied to their control windings.

12·15. Hydraulic and Pneumatic Devices.—The possibility of using hydraulic, electro hydraulic, pneumatic, or electro pneumatic power devices will often arise in the survey of a particular servo problem. The choice among these devices and electronic or electrical devices will, in

[1] L. Andrews and F. Shanely, "One Type of Rotary Magnetic Clutch and Its Associated Brake Used on Aircraft Electric Motors," *Elec. Eng.*, **63**, 893, December 1944.

[2] W. H. Gille and H. T. Sparrow, "Electronic Auto-pilot Circuits," *Electronics*, **17**, 110, October 1944.

general, be based both on a comparison of the devices considered themselves and on the facilities and experience of the designer and producer.

Hydraulic and pneumatic power devices have the advantage of fast response; in this respect they can be made to surpass the best electric motor drives and are about on a par with the best magnetic clutch systems. In hydraulic systems, for example, accelerations of 100,000 radians per second-per-second and velocity errors of $\frac{1}{350}°$ per degree-per-second

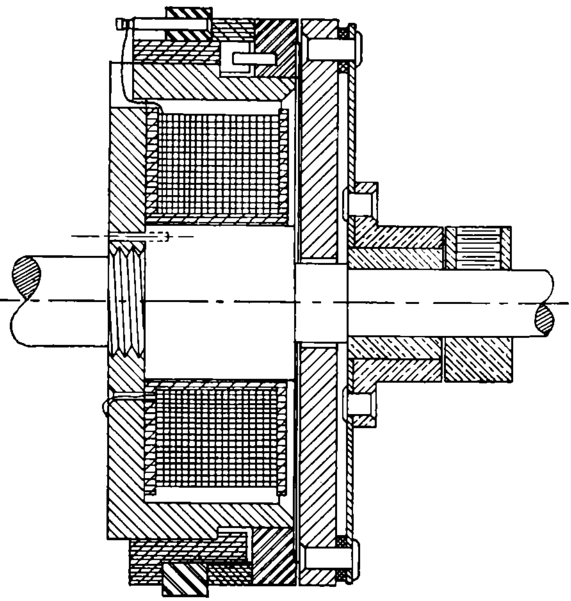


FIG. 12·23.—Moving-coil friction-type magnetic clutch.

have been reported.[1] Conventional hydraulic servos may achieve accelerations in excess of 20,000 radians per second-per-second. Pneumatic devices are simple and rugged and in general may be produced slightly cheaper than competing electronic equipment when large quantity production is concerned. When properly designed, little maintenance is required for these devices. For these reasons, and also because to some extent some of the techniques of electronic control have been newer developments, hydraulic and pneumatic systems have been extensively used for industrial process control, where they are used in systems for control of temperature, flow, level, and many other similar applications. Pneumatic servos have been used occasionally in airborne applications; the T-1 bombsight computer and the German V-1 control mechanism are

[1] "Fundamental Studies in Servomechanisms Rated Approximately 100 Watts," MIT Servomechanisms Laboratory, September 1943.

typical examples. One of the most serious objections to hydraulic and pneumatic systems is the problem of piping control or data leads long distances. Such an installation is inflexible; and if damaged, the entire sealed system may have to be removed for repair. Long piped leads also introduce time lags which at best are undesirable and at worst may seriously interfere with the response of a control system. Hydraulic systems are usually used for heavy loads, and pneumatic systems for the lighter loads. Pneumatic systems offer important anti-explosion features.

Introductions to the conventional techniques of process control are found in two recent volumes by Eckman[1] and Smith.[2] The first gives a good treatment of the characteristics of many control elements and includes a fairly complete bibliography. The latter includes appendices reviewing the mathematics of transient analysis and contains an extensive anotated bibliography. These books are written primarily from the hydraulic and pneumatic controller point of view and use many terms that will sound unfamiliar to the electronic circuit designer; in the light of available electronic developments some of what is described seems complicated or cumbersome. However, the methods presented represent the techniques of a large and important industry, with a long record of dependable performance, and should not be lightly regarded.

When hydraulic and pneumatic devices are considered from the standpoint of their relation to the designer and producer, other factors become important. The first of these is the relative unavailability of ready-made, instrument-size hydraulic devices. To some extent, the availability problem is less acute in the pneumatic device field. Where relatively small numbers of instrument servos are to be used, design cost and time become relatively important and will usually be higher per unit for hydraulic or pneumatic than for electronic devices. The principal reason for this is the availability and cheapness of standard electronic devices such as motors and vacuum tubes and the ease with which they can be adapted to new applications. The production facilities required for producing electronic servos are generally less critical than those for producing hydraulic or pneumatic servos. In summary, it may be said that hydraulic and pneumatic devices offer rapid response and dependability, and, for some applications, economy, but may be more difficult to design, produce, and maintain. Their advantages for many applications are sufficiently great, however, to offer a real challenge to the designer of electronic equipment.

A recent *Fortune*[3] magazine article on the process control instrument industry examines one company in detail and presents an interesting

[1] D. P. Eckman, *Principles of Industrial Process Control*," Wiley, New York, 1945.
[2] E. S. Smith, *Automatic Control Engineering*, McGraw-Hill, New York, 1945.
[3] "Taylor Instruments," *Fortune*, **34**, No. 2, 89.

picture of the competitive conditions in the field and is recommended for further reading.

12·16. Control of D-c Motors.—The purpose of the power-control circuit of a servomechanism is the utilization of information in the error signal and its derivatives and integrals to control the sense and the amount of the power output of the motor in such a manner as to reduce the error and hence the error signal. The control circuit itself may be as simple as a switch or relay, connecting the motor to a power main in the correct sense, or it may be a more complicated device such as a thyratron or hard tube circuit or a still more complicated device such as an Amplidyne or other controlled d-c generator.

Relays have the advantages of being light, compact, and simple. The disadvantages of ordinary relays are short life and limited load-handling capacity. Special relays,[1] such as the Western Electric D-168479 relay with mercury-wetted contacts, have successfully stood up under severe service and are satisfactory for many applications, although they cannot be operated in all positions. By the use of suitable "buzz" or "flutter" circuits, relay action may be made to approach continuous control, in that power applied to the motor may be varied smoothly from maximum in one direction through zero to maximum in the opposite direction. In relay servos, great care must be exercised in the design of suitable networks for contact protection; lack of adequate networks may result in short relay life. The principal reason for this is the high inductive voltage that results when motor current is interrupted by an opening relay contact.

Thyratrons are slightly lighter than relays but cannot control as much current. Small thyratrons, such as 2050's, can handle hundreds of milliamperes, whereas mercury-contact relays can handle amperes. Thyratrons are not so suitable as relays for control of motors directly from low-voltage supplies because of the voltage drop across the thyratron. A thyratron motor circuit usually involves either a transformer, which constitutes a major part of the weight and the bulk of the circuit, or a d-c supply, which is even heavier and bulkier.

Vacuum-tube drive gives the smoothest control of any type so far mentioned, but with proper design other types of drive circuits are usually also satisfactory in this regard. In the case of field-controlled motors and other devices that because of their inherent power amplification may be directly controlled with small vacuum tubes this type of control is usually the lightest and simplest. When armature currents must be controlled, vacuum-tube control becomes comparable in weight and complexity to thyratron control but is somewhat less efficient.

[1] See Sec. 12·20.

Controlled generators such as Amplidynes or Ward-Leonard generators have been used extensively for higher-power servos; however, for the instrument servo field their weight and complexity compare unfavorably with the simpler power-control methods discussed.

It is often important in choosing or designing a power-control circuit that the motor be operated at high efficiency, for with any given motor and load conditions, the maximum load acceleration is directly proportional to the torque output of the motor. A change of amplifier gain alone, while increasing the gain of the system, will not increase the maximum torque available at the motor but will merely allow the system to overload at smaller values of error, the resulting maximum acceleration remaining unchanged. Consequently, for maximum performance, every precaution should be taken to permit high power input to the motor and high motor efficiency. Specifically, this means that the motor should be designed for low core losses, should have an allowable temperature rise as large as possible, and should be mounted to dissipate as much heat as possible by conduction, radiation, and convection or forced circulation. For most motors, an extremely important factor in motor efficiency is the waveform of the control voltage applied to the motor, that is, the relative amounts of a-c and d-c power applied to the motor. Since only the d-c power can produce useful work in armature-controlled or field-controlled d-c motors, the percentage of a-c power delivered to the motor should be kept as low as possible with these types of motors.

It is usually desirable to reduce power at the servo balance or null position to the point where either a-c ripple peaks or imposed signal peaks provide just enough torque to vibrate the armature, which means enough to overcome brush, bearing, gear, and load friction and load inertia. A preferred type of operation involves alternate short pulses of full motor power in opposite directions separated by periods of zero power. This type of operation is more efficient than the type where full power is always applied in either one direction or the other, but the latter type is usually easier to provide and is therefore more common. Either type greatly reduces the effect of starting friction and hence results in a system that is very sensitive to small signals.

Comparison of Full-wave and Half-wave Power-control Circuits.—The action of thyratrons and vacuum tubes used to control d-c motors from a-c power sources may be compared to the use of controlled rectifiers. Conventional controlled rectifier theory applies to it with only a few special provisions.[1]

[1] P. E. Mayer, "Electronic Control of D-c Motors," *Electronics*, I, 98, May 1943; II, 119, June 1943; III, 118, July 1943; IV, 133, September 1943; V, 128, October 1943; P. T. Chin and G. E. Walter, "Transient Response of Controlled Rectifier Circuits," *AIEE Trans.*, **64**, 208–214, April 1945; P. T. Chin and E. E. Moyer, "A

Factors that must be considered in evaluating the voltage and current waveforms applied to the motor are a-c supply voltage, inductance and resistance of motor, back emf of motor, line frequency, voltage drop across control tube, and firing angle in the case of thyratrons.

When a half-wave rectifier such as a grid-controlled thyratron is used without a large capacitance to supply power to a d-c motor, continuous current flow cannot be achieved, regardless of the combination of values, except during transients when a motor emf of polarity such as to increase the flow of current is present or if a backswing damper is used. A large capacitance, while permitting continuous current flow, draws large currents for short periods from the rectifier. A backswing damper is an element possessing rectifier characteristic; it is connected in shunt with the motor load so that when a positive voltage is applied to the motor load through the controlled rectifier, the damper appears as a large resistance connected across the motor load. When rectifier current is interrupted, motor current may continue to flow in the circuit formed by the motor and the backswing damper in series. Backswing dampers have been found to be very useful and efficient in the unidirectional control of d-c motors from half-wave rectifier circuits. Where bidirectional control of motors is required, backswing dampers cannot be used across any element in which the direction of current flow is reversed in order to achieve a reversal of direction of rotation of the motor. However, they can be used across the two halves of a split motor field where each half of the field carries current in only one direction. Through their use, the power that can be transferred to a motor by half-wave drives is greatly increased. An increase of a-c frequency for conditions where the current flow is not continuous and all other parameters are kept constant results in a decrease of the power that can be transferred to the motor load. Where current flow is continuous, change of frequency affects principally the percentage ripple, but not the average current. If not obvious, the explanation of these statements may be found in the paper of Chin and Moyer previously cited.

Full-wave control at low firing angles has much the same characteristics as half-wave control. Current is not continuous, and power transferred to the load decreases with increasing frequency. A backswing damper may be used to get continuous flow of current where discontinuous flow would otherwise obtain. As firing angle is increased, a point is suddenly reached at which current flow becomes continuous. Chin and Moyer point out that current is discontinuous as long as the conduction of one tube in a full-wave circuit terminates before the start of

Graphical Analysis of the Voltage and Current Waveform of Controlled Rectifier Circuits," *AIEE Trans.*, **63**, 501–508, July 1944.

conduction in the other tube. When termination of current flow in one tube coincides with initiation of current flow in the other tube, continuous current flow prevails.

The build-up of current to its steady-state condition starting from zero or discontinuous flow of current has been treated by Chin and Walter.[1] Under continuous flow conditions, the average current flowing in the load does not change with frequency. However, the percentage ripple does, ripple power decreasing as the frequency is increased. In full-wave control circuits, the transition from discontinuous to continuous current flow as the firing angle is changed may result in an abrupt change of slope in the motor torque vs. control signal characteristics of the combination of motor and control circuit. When current flow is continuous, the presence of a backswing damper does no harm and may increase slightly the average current that flows, since the voltage applied to the motor is forced to remain positive.

By way of illustrating the magnitudes of these effects, it has been found advisable to derate d-c motors used in discontinuous current circuits by a factor of roughly 50 per cent, and those used in continuous current flow circuits by roughly 25 per cent of their normal d-c ratings.

Limit Switches.—Some means is usually needed for stopping the motor of a positioning servo at the desired travel limits. This may take any of several forms but usually involves the use of limit switches operating at the desired limit of travel. When operated, the limit switch removes driving power to prevent further travel but leaves the drive circuit for reverse travel connected. If no other precaution is taken, overtravel even with limit switches may be excessive.

To prevent this, some form of dynamic braking may be obtained by the limit switch connection and perhaps by auxiliary equipment. A single-pole double-throw limit switch can be connected to remove drive power and at the same time to connect a suitable load across the motor for dynamic braking. For armature control, as in the circuit of Fig. 12·30, a rectifier of suitable capacity can be connected as a dynamic-braking load without changing the drive conditions for the reverse direction.

Alternatively, the drive circuit itself may be reconnected by the limit switch so as to apply reverse power to prevent or minimize overtravel. With such an arrangement, the servo will then drive away from the limit switch until the switch is released, with the result that the servo will oscillate about the limit switch position as long as the servo input corresponds to an output position beyond the limit. With such oscillation, power dissipation in the motor will be high and may cause excessive heating and excessive wear on the gear train. This method is therefore

[1] *Loc. cit.*

recommended only for applications where limit input signals are usually avoided or where operation at the limit switch is required for short periods of time only.

Practical limit switches are characterized by an appreciable operating differential travel. This is required in order to get snap-action interruption of the circuit. The fraction of the servo scale that is represented by the differential travel is dependent on the type of actuator used and the mechanical design of the gear train as well as the particular switch chosen. The limit-switch operating mechanism should, of course, be designed to minimize this operating differential. For higher power drives some contact protection may be required even with snap-action switches. The principles of relay contact protection outlined in Sec. 13·20 apply here. Several companies make switches suitable for servo limit switching.[1]

One other method of limit protection is important. A combination of mechanical limit stops and a slip clutch can be used either alone or with limit switches to remove motor power. A slip clutch may be inserted between the motor and the servo output shaft, because the motor may slip any number of revolutions without affecting servo operation.

The output shaft is decelerated by the mechanical limit stop, while the slip clutch, stopped on the output side by this limit stop, applies a decelerating torque to the armature, bringing it gradually to a stop. The design of slip clutches is discussed in Vol. 17. This type of limit protection may be used in small units without disconnecting the motor power at the limit of travel. Because of the low power dissipation and short life of small clutches while slipping, however, their use is best restricted to very low power applications or to those applications where input signal values requiring frequent slippage are not encountered. In other applications, limit switches should be employed to remove motor power at the limit of travel.

One interesting variation of a clutch arrangement is employed in certain Lear motors.[2] Here motor power is applied through a magnetic clutch energized in series with the motor windings. When power is removed by operation of the limit switch, the driven clutch plate is spring-loaded in the unactivated direction to bear against a stationary friction surface, thus acting as a friction brake on the driven gear train to minimize overrun.

[1] The list includes the Microswitch Co., the Mu-Switch Co., and the Acro Electric Co. A particularly useful switch is the General Electric Co. Switchette which is extremely small in size, has a high current rating and a satisfactorily long life. A variety of contact arrangements can be obtained from all of the manufacturers listed. See Vol. 17 of this series for further details.

[2] L. Andrews and F. Shanely, "One Type of Rotary Magnetic Clutch and Its Associated Brake Used on Aircraft Electric Motors," *AIEE Trans.* **63**, 893, December 1944

12·17. Thyratron Control.[1]—Grid-controlled, gas-filled thyratron rectifier tubes offer attractive possibilities for d-c motor control. Recent developments have produced several inert gas-filled types that are relatively insensitive to temperature or mounting position and have average current ratings in the range from 75 ma to several amperes.

Thyratrons permit some economy of circuit design, power consumption, size, and weight. They require a few special circuits not needed for hard tubes; but if the circuits are properly engineered, they are capable of reasonably smooth, proportional control and are adaptable to a wide variety of operating conditions.

Thyratron Characteristics.—A thyratron behaves essentially like a switch in series with a small direct voltage. The switch is closed (i.e., the tube fires) if, for a given value of positive anode voltage, the grid potential exceeds a critical value. Current, for values less than the peak emission of which the cathode is capable, is limited by the anode circuit load and the applied voltage minus the internal tube drop. The tube will not fire for any grid voltage until the anode voltage exceeds a certain minimum value; after it fires, the anode-to-cathode voltage drops to a low value—approximately 15 volts for mercury tubes and between 8 and 20 volts for other gas tubes. Once the tube fires, the grid normally loses control until the anode potential falls below an extinction value that depends on the type of tube and the grid voltage but is usually of the order of 25 volts. An exception to the statement that the grid loses control must be made for special conditions where high negative grid voltages can interrupt low anode currents. This phenomenon, however, is of no practical use for motor control.

A finite time is required for deionization before the grid can regain control after the anode potential drops to zero. This may vary from 20 to 100 μsec for the inert-gas tubes and from 100 to as high as 1000 μsec for mercury vapor tubes. This deionization time increases with increasing anode current and, for mercury tubes, with increase of bulb temperature. At commercial power frequencies no deionization trouble is usually encountered with either type tube, except in some special applications such as inverters. For mechanical and thermal reasons, the mercury tubes are not suitable for aircraft usage. Inert gas thyratrons have been used satisfactorily in airborne equipment.

It is desirable to delay the application of anode voltage until the cathode has been heated to the point where emission is higher than needed for the circuit conditions. If tube drop exceeds a critical value in the neighborhood of 20 to 25 volts for mercury vapor or inert gases, as may happen if anode voltage is applied while emission is low or if the maximum safe current is exceeded, the cathode will be bombarded by positive ions

[1] Sections 12·17 to 12·19, inclusive, are by J. R. Rogers.

of sufficiently high energy to cause disintegration of coated-type cathodes. At normal tube drops, the ions have insufficient energy to cause appreciable damage. Accordingly, time delay relays or manual sequence controls are necessary if maximum tube life is to be obtained.

For small servos using type 2050 or similar thyratrons it is often possible to secure adequate tube life without time delay relays, particularly where the equipment is operated for relatively long periods without being turned off. In some cases the operation of other equipment such as thermionic voltage regulators may involve suitable delays for the operation of thyratron plate power relays. Alternatively the heater warm-up delay of a tube controlling a plate power relay may be used, with the tube heater operated in parallel with the heaters of the thyratrons. Such thermally-controlled power circuit relays are better than nonthermal relays in that if power is momentarily interrupted, operation can be restored. However, variability and aging of tubes makes the circuit difficult to control accurately.

Despite these special characteristics, thyratrons offer for many applications the most economical control method. Low tube drop means that low transformer voltages may be used. A relatively low-impedance (hence higher current) motor can be used without excessive power loss in the rectifier. Low-impedance motors are easier to wind, are less apt to fail because of insulation breakdown, and are more likely to be commercially available. Any increase that thyratron control may produce in drive efficiency is, of course, most significant under conditions where the power to be controlled is large relative to the fixed heater and amplifier power required. For example, thyratrons would hardly be enough more efficient than vacuum tubes for relay operation to justify their increased cost on the basis of efficiency alone.

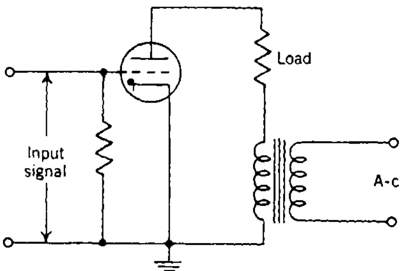


FIG. 12·24.—Simple thyratron control circuit.

The nature of thyratron tube characteristics does require that certain simple circuit precautions be observed if smooth motor control comparable to the best other control methods is to be attained.

By way of illustration, control with the simple thyratron circuit of Fig. 12·24 is of an off-on nature. If the voltage applied to the grid of the thyratron exceeds, at any time during the half-cycle of positive anode voltage, a critical value dependent upon the instantaneous anode voltage, the tube fires and conducts for the balance of that half cycle, extinguishing when the anode voltage drops below a value of approximately 20 to

30 volts determined by tube design. During the period of conduction, the internal tube drop is nearly constant at a value determined by such design factors as the nature of gases and gas pressure within the tube. Figure 12·25 shows the relation between anode voltage and critical grid potential for operation from an a-c anode supply. Referring to Fig. 12·25*b*, it will be seen that a change of d-c grid bias from the value *A*, at which the tube never fires, to the value *B* will result in the tube firing once each cycle at the time when the critical firing voltage becomes just equal to the grid bias. In this case, the tube will fire for the second half of each positive half cycle. For a new and less negative value of grid-bias voltage *B′* this point of intersection of the bias with the critical firing voltage will occur earlier in the half cycle; and by proper choice of bias, any value of current flow into the load can be obtained between half (90°) and essentially the entire positive half cycle (180°) of anode supply voltage. Angles of current flow between 0° and 90° cannot be secured with this arrangement, because a d-c bias cannot first intersect the critical firing curve later in the positive part of the cycle than the point of maximum anode voltage. Power supplied to the load, therefore, may be varied smoothly by varying the d-c grid bias from half to full power; but when the negative bias is increased, the power delivered jumps suddenly from half power to zero power. Such an arrangement used for servo motor control would seriously increase the difficulties of stabilization, because the system would have essentially an off-on type of control over part of its range.

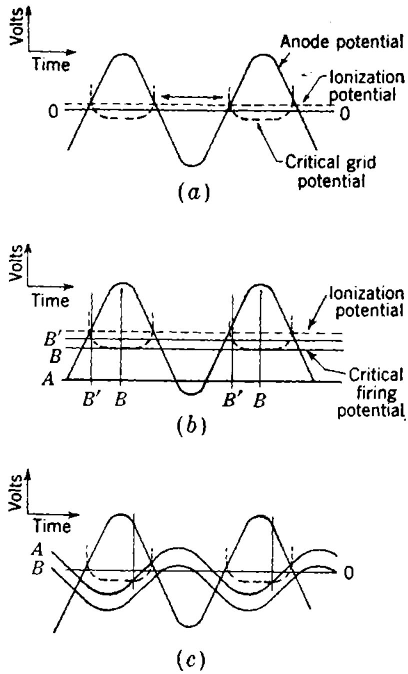


FIG. 12·25.—Firing characteristics of thyratron circuits. (*a*) Relation between anode potential and critical grid potential; (*b*) effects of change of grid potential on firing point; (*c*) control of firing point with d-c plus lagging-phase a-c bias.

To avoid these difficulties, an a-c bias is usually superposed on the d-c grid bias. The special case of interest here is that of an a-c grid bias voltage that lags the anode voltage by an angle between 90° and 150°. Figure 12 25*c* illustrates the control obtained with a combination of d-c bias and a-c lagging phase bias. At B no conduction occurs; at A, a less negative d-c bias, the critical bias is crossed near the end of the cycle, the tube fires, and current flows for a negligibly small angle. For d-c bias values increasingly more positive than A, the angle of flow increases smoothly until conduction occurs over essentially the entire half cycle. A square-wave voltage in phase with the anode voltage and superimposed on the combination d-c and a-c bias A of Fig. 12·25c could likewise produce

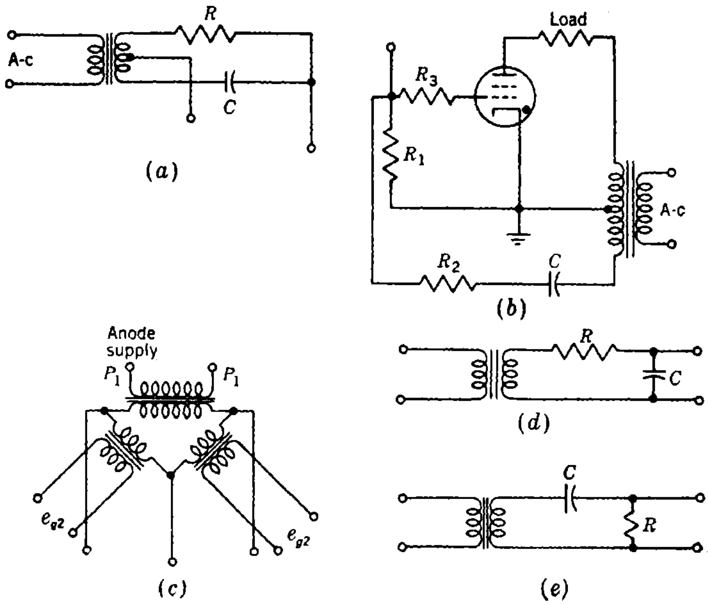


FIG. 12·26. —Practical circuits for securing phase-shifted a-c bias.

smooth control proportional to the amplitude of the square wave. Effectively, the same result can be secured with a sine wave error signal with amplitude proportional to the error and in phase with the anode voltage. In this case, the curve of firing angle against input signal amplitude is obviously altered, but smooth control is secured because, in effect, the phase of the net a-c grid voltage is advanced and the net bias is made to cross the critical value earlier in the cycle as the control voltage is increased.

Practical circuits for securing phase-shifted a-c bias are shown in Fig. 12·26. The choice of circuits depends, in some measure, on the control circuit, on the available transformers, and on the nature of the power supply. In some cases, it will be necessary or desirable to introduce some filtering in the supply to the phase-shifted output to reduce the effects of harmonics. Circuit 12·26d is better than 12·26e in this respect, employing the drop across C as the phase-shifted output. This lags the voltage applied across R and C. For pure sine wave input and a 90° phase shift, taking the voltage across R and reversing the phase of the input will be equivalent; if harmonic content is high, this will result in higher harmonic content of the output.

Various other methods of controlling the thyratron firing angle have been proposed. One of interest[1] consists of developing a pulse whose time of occurrence relative to the anode supply waveform is determined by a voltage-sensitive time-modulation circuit such as a cathode-coupled multivibrator (*cf.* Vol. 20). The output of such a multivibrator may be used to trigger a pulse generator, this pulse being used to fire a normally nonconducting thyratron. For some applications such a circuit may be useful; usually it will be too complicated.

The firing angle may also be controlled by phase shifting. This phase shift can then be controlled by the output of the servoamplifier, often with an appreciable gain in power. For example, direct current driving a saturable reactor may be used to provide phase control of the net grid bias. Likewise d-c bias control of the gain of an amplifier tube in a "fed-back time constant" or Miller effect circuit may be used. The equivalent C thus obtained is used as part of a fixed R, variable C phase shifter. Difficulty will usually be found in securing a smooth control over the range from zero to full half-cycle conduction with devices of this type, and they are not in general recommended.

It is extremely important, if smooth and symmetrical control is to be secured, to eliminate the pickup of stray a-c voltages. One of the commonest sources of trouble in this regard is the fact that power transformers of conventional design rarely have adequate interwinding shields, and in particular, they usually have no shields between the high-voltage windings and the outside-wound heater windings. If a heater winding of such a transformer is used as a supply source for a phase-shifter circuit, enough alternating current may be coupled into the phase-shifter circuit through the interwinding capacitance to mask completely the desired phase-shifted voltage. Trouble from this source will increase as frequency of line power increases and may be serious if the a-c power supply has a high content of second and higher harmonics for which the interwinding capacitance represents an impedance lower than for the fundamental. A phase-shifter circuit that can be grounded or run at low impedance to ground is thus preferred. The phase shift for any harmonics present will be different from that for the fundamental and may be different for the different drive tubes.

Several types of thyratron tubes are available, differing with regard

[1] M. M. Morack, *Gen. Elec. Rev.*, **37**, 288 (1934); H. J. Reich, *Theory and Applications of Electron Tubes*, McGraw-Hill, 1939, Sec. 12–32.

to temperature sensitivity, control-grid characteristics, and suitability for mobile or portable applications. For average currents up to approximately 5 amp a wide variety of tubes filled with inert gases are available. These tubes, filled with argon, krypton, xenon, or a mixture of these gases, are relatively insensitive to bulb temperature, mounting position, and vibration, as compared with mercury-vapor tubes. The latter are available with average current ratings from 0.25 amp upward, the largest sizes approaching in capacity the pool-type tubes of the ignitron class.

For instrument servo applications only tubes with ratings below 1- or 2-amp average current are of interest. Table 12·2 summarizes the most commonly available units.

TABLE 12·2.—IMPORTANT CHARACTERISTICS OF THYRATRON TUBES SUITABLE FOR OPERATION OF LOW-POWER SERVOS

Tube type	Mfr.	Max. anode voltage (forward), volts	Peak anode current (max.)	Average anode current* (max.)	Inert gas or mercury vapor	Triode or shield grid	Heater or filament	Heater power
GL.546	GE		100 ma	10 ma	Gas	Shield grid	Heater	6.3v 0.2 amp
2D21	RCA	650	500 ma	100 ma	Gas	Shield grid	Heater	6.3v 0.6 amp
884	RCA	350	300 ma	75 ma	Gas	Triode	Heater	6.3v 0.6 amp
2051	RCA	350	375 ma	75 ma	Gas	Shield grid	Heater	6.3v 0.6 amp
2050	RCA	650	500 ma	100 ma†	Gas	Shield grid.	Heater	6.3v 0.6 amp
3D22	RCA	650	6.0 amp	.75 amp	Gas	Shield grid	Heater	6.3v 2.6 amp
FG-17	GE	2500	2.0 amp	0.5 amp	Vapor	Triode	Filament	2.5v 5.0 amp
FG-27A	GE	1000	10.0 amp	2.5 amp	Vapor	Triode	Filament	5.0v 4.5 amp
FG-33	GE	1000	15.0 amp	2.5 amp	Vapor	Triode (pos. grid)	Heater	5.0v 4.5 amp
FG-97	GE	1000	2.0 amp	0.5 amp	Vapor	Shield grid	Filament	2.5v 5.0 amp
FG-98-A	GE	180	2.0 amp	0.5 amp	Gas	Shield grid	Filament	2.5v 5.0 amp
FG-178A	GE	310	0.5 amp	0.125 amp	Gas	Triode	Filament	2.5v 2.25 amp
ELC1A	Electrons, Inc.	170	2.5 amp	0.40 amp	Gas	Triode	Filament	2.5v 6.0 amp

* Averaged over 30 sec.
† Higher currents allowed for operation at low anode potentials.

Two basic tube constructions, triodes and tetrodes, are available both in inert gas and mercury-filled tubes. With the straight triode thyratrons (types 884 and FG-67, for example), a relatively high grid current flows even before the tube fires. If such a tube is operated from a high-impedance circuit, this current may cause sufficiently high voltage drops in the input circuit to interfere seriously with the intended operating conditions. Because grid current will flow anyway after the tube fires, a protective resistor must be used in series with the grid to limit the grid current. If the value of resistance used is too large, the drop across the resistor caused by grid current in the nonconducting state may even prevent firing.

The shield grid thyratron (types 2050, 3D22, and FG-98, for example), was designed to minimize this grid current previous to firing. Such tubes have a somewhat different control characteristic and a relatively high grid impedance. It is possible to use a grid circuit resistance as high as 10 megohms with the 2050 tube, whereas the rated maximum value for the corresponding triode, type 884, is $\frac{1}{2}$ megohm.

Mercury-filled tubes will operate satisfactorily only over a rather limited range of temperature because of the rapid change with temperature of the vapor pressure of mercury in a saturated system. Tubes filled with a mixture of mercury and one of the inert gases are somewhat better in this respect. Where operation over a wide range of bulb temperature is necessary, tubes filled with inert gases are preferred. Most inert gas tubes now available can be operated between $-50°$ and $+90°C$.

A further advantage of the inert gas-filled tubes is that having no liquid mercury, they are not critical as to operating position and only moderately sensitive to vibration, but no more so than vacuum tubes with similar mechanical construction. For instrument servo control, gas-filled thyratrons of the shield grid type are therefore recommended.

One consequence of the finite grid circuit impedance of thyratrons is that where two tubes are driven from a single source of moderately high impedance and appreciable reactance, such as a coupling condenser, it is possible for grid current of one tube to cause a bias to build up that tends to prevent the other tube from firing at its normal control point. This effect (or similar effects caused by feedback from the fired tube to the amplifier input) gives rise to a peculiar servo performance in which the apparent balance point is different for the two directions of rotation. As the error signal is reduced, the servo approaches a balance point that turns out to correspond to an overshoot and gives a reverse error. In driving back in the opposite direction a similar effect is present. Under these conditions, a control circuit and motor may hunt with an amplitude that cannot be controlled by varying the amplifier gain. The remedy is to isolate the grid circuits of the thyratrons for both directions of drive, to reduce common impedances, and to increase the individual series current-limiting resistors to reduce grid current. In some cases it may be desirable to drive the tubes for each direction from a separate isolating tube or otherwise isolate the control. For example, the circuit of Fig. 12·27 has separate d-c grid returns for the two tubes. Further isolation can be had by increasing the series grid resistors from 100 to 200 k or even higher, by lowering the output impedance of the driving amplifier, or by using a final amplifier stage consisting of two tubes operated with grids in parallel but with separate plate circuits driving each thyratron from the plate of one of these tubes. The two-null condition described is not a frequent occurrence, but it can be very confusing when encountered unless the designer is prepared for it. It may also occur with other types of servo motor controls but has been discussed here because it is more apt to occur with thyratron controls.

For the reasons discussed in Sec. 12·16, the use of full-wave drive is to be preferred where variable-frequency power supply is a requirement. Operation over a range of frequency of about 2 to 1 is possible if the phase-shift circuit is correctly chosen. For a much wider range of frequency, proportional control may be lost at the extreme supply frequencies. This may prevent proper stabilization.

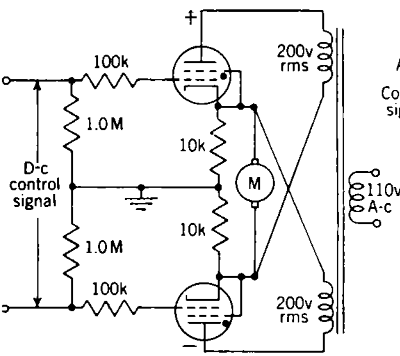


FIG. 12·27.—Circuit for thyratron bi-directional control of motor armature from d-c error signal.

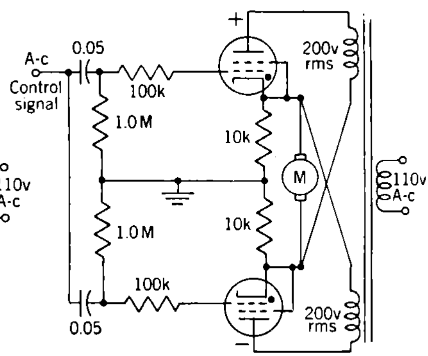


FIG. 12·28.—Circuit of Fig. 12·27 used with a-c error signal.

The problem of operation with high supply frequencies is important for airborne equipment because of the power supply economies at high frequencies and particularly when engine-driven variable-frequency generators can be used. To date, the tendency among designers has been to avoid the variable-frequency problem, if possible, by using 400 cps or lower power frequencies or, if necessary, by using other types of control (such as relay control), obtaining motor power from the 28-volt d-c supply of the aircraft. A compact a-c operated servomechanism that is independent of power-supply frequency would help considerably in simplifying aircraft control and navigational equipment.

It may be desirable for some applications to operate a motor directly from a high-voltage (110 to 220 volts or higher), d-c supply. Thyratrons are desirable for control tubes in such an application because of their low internal voltage drop and high current-carrying capacity. Because the thyratron grid normally loses control over plate current when the tube fires, special provision must be made to reduce the anode potential when it is desired to extinguish the tube. In practice, the use of two thyratrons in an inverter circuit offers the best approach to this problem.[1] Basically such circuits operate so that when the nonconducting tube fires, it causes the anode potential of the conducting tube to drop below

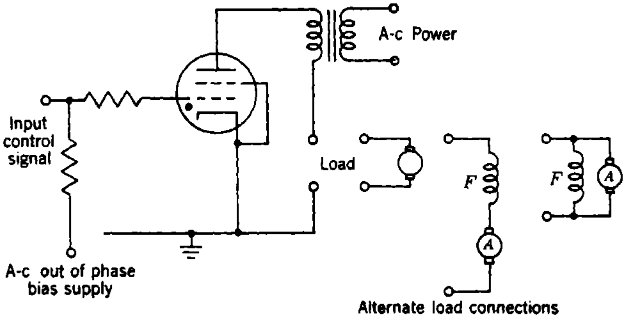


FIG. 12·29.—A basic thyratron motor-control circuit.

the extinction point and causes a transfer of load current to the first tube. Interconnection with suitable time constants makes the process automatically repetitive. By injecting control signals of suitable nature and by use of auxiliary control tubes, the action of the inverter may be controlled to vary the load power.

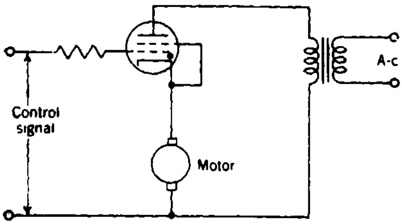


FIG. 12·30.—Basic motor-control circuit with cathode load.

Such devices are primarily suitable for somewhat higher power applications than those of this chapter. For most instrument servos the use of vacuum tubes, grid controlled by a d-c error voltage, will be the best way to operate a motor directly from a high-voltage d-c supply. For low-voltage supplies, direct relay control may be used to advantage.

12·18. Practical Thyratron Motor-control Circuits.—Figures 12·29 and 12·30 show the thyratron circuits on which are based most of the other circuits given. A relatively small change of grid voltage is sufficient to cover the full range of motor power for normal values of lagging a-c grid bias. Motor speed is controlled principally by the transformer voltage, generated back emf, load friction, and the friction losses in the motor. If the motor load is put in the cathode circuit of the thyratron, a higher degree of speed regulation is obtained, particularly if the load consists of only the motor armature. The point in the a-c cycle at which the

[1] See, for example, Reich, *op. cit.*, Sec. 12–44.

thyratron fires will depend on the d-c grid-to-cathode voltage while the tube is nonconducting; and with the motor in the thyratron cathode circuit, this depends directly on the generated emf, a quantity approximately proportional to speed. Hence, the motor speed will be almost proportional to the error signal. This characteristic is highly desirable from the standpoint of both servo stabilization and the linearization of motor characteristics that it provides but does require high gain and swing in the error signal amplifier.

Figure 12·31 illustrates two adaptations of the basic circuit of Fig. 12·29, designed for half-wave armature control of a separately excited field motor with relatively high-impedance armature operated at 100 to 115 volts direct current. The circuits of Fig. 12·31 float at the out-

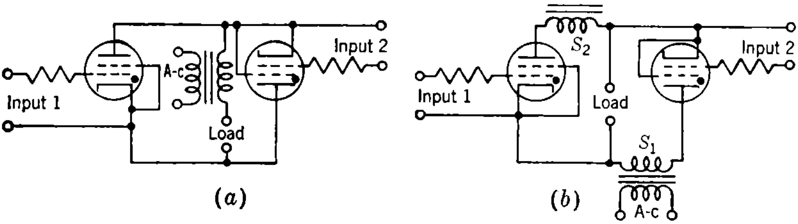


FIG. 12·31.—Bidirectional drive circuits.

put potential of the d-c error amplifier, while that of Fig. 12·27 can be tied down by grounding the center of the resistor divider across the motor. The outstanding characteristic of circuit of Fig. 12·27 is inherent speed control obtained by cathode feedback of half the armature voltage. If lagging grid bias is used, motor speed will be roughly proportional to error signal up to high speeds. This feature together with rapid response to changes in the error signal makes the circuit capable of surprisingly smooth and linear servo performance with a minimum of additional stabilization and goes far to make up for the disadvantages of half-wave drive. Reversal of rotation can be extremely rapid with this circuit, as armature voltage adds to the transformer voltage on reversal, providing extra current for stopping and reversal. Power output is about 0.003 hp for the Elinco[1] Midget B-35 or PM-1-M permanent-magnet field motor and 0.005 hp for the Elinco B-64 motor which has a lower-impedance armature and a 28-volt wound field. These figures are for a 400-cycle design; higher outputs can be secured at 60 cycles. The input signal for maximum speed is high, from 125 to 150 volts d-c grid-to-grid, but this value can be obtained with 6SL7 or 6SN7 amplifiers and offers no serious difficulties.

A circuit for bidirectional full-wave thyratron control of split-series

[1] Electric Indicator Co., Stamford, Conn.

motors is shown in Fig. 12·32. Care must be taken to prevent both sets of drive tubes from firing together, as the resultant field flux in the split-series fields will then cancel and cause the motor torque to fall to zero. This circuit has inherent speed control characteristics because of the motor armature voltage feedback to the grid circuit. This contributes to the stability of this servo.

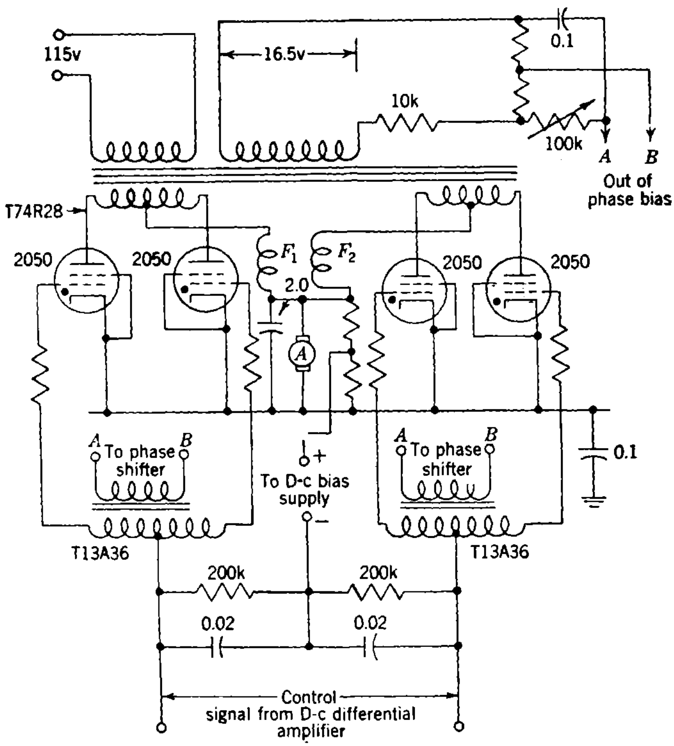


FIG. 12·32.—Push-pull bidirectional thyratron control circuit for Elinco MS-2 or FB split-series motor.

The chief disadvantage of the circuit of Fig. 12·32 is the fact that it floats above ground at a potential determined by the amplifier output level. The motor must therefore be insulated for approximately 200 volts, and insulation failure will disrupt the drive circuit operation. This difficulty can be avoided either by use of a d-c amplifier with an output at ground potential in the absence of a signal or by the use of a-c amplifiers driving the thyratrons. With either of these arrangements, the thyratron cathodes and one end of the armature can be grounded.

Direct current amplifier output at or near ground potential requires a negative power supply and may require an appreciable gain reduction.

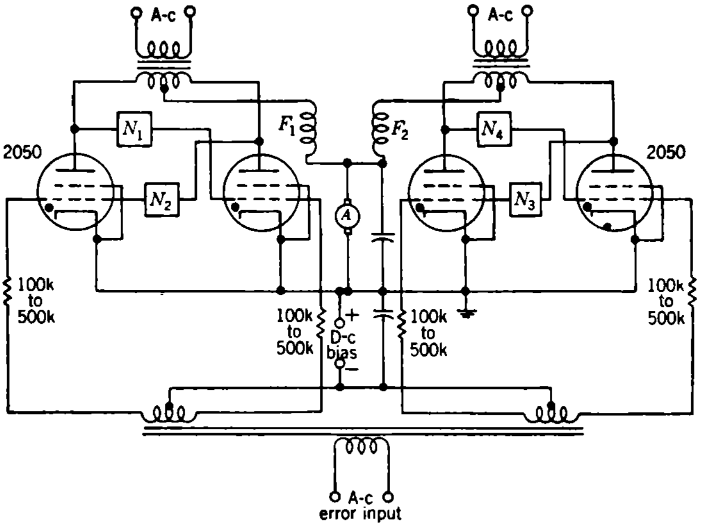


FIG. 12·33.—Bidirectional push-pull thyratron motor-control circuit with a-c error signal for split-series motor. Out-of-phase bias is supplied by networks N_1, N_2, N_3, and N_4.

Level changing circuits are discussed briefly in Chap. 3. The use of a-c error signals eliminates the problem of d-c amplifier drift and by the use of suitable transformers may be adapted to most drive circuits. Figure 12·33 illustrates a-c drive to a push-pull circuit for split-series motor control. The use of transformers for both the error signal and the out-of-phase grid bias is cumbersome, and it is preferable to derive this bias directly from the high-voltage plate transformer through the use of a phase-shifting network.

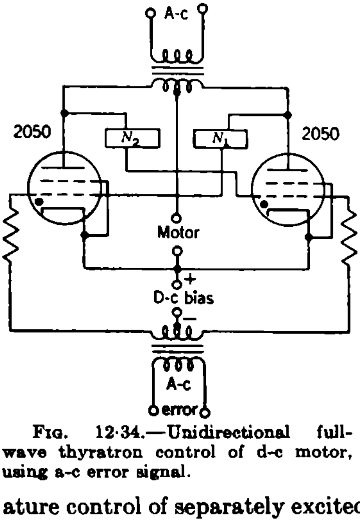


FIG. 12·34.—Unidirectional full-wave thyratron control of d-c motor, using a-c error signal.

Where unidirectional operation is desired, or where the motor connections can be satisfactorily reversed by a switch or relay, the push-pull circuit of Fig. 12·34 is useful. It is suitable for control of series motors or for armature control of separately excited field or permanent-magnet field motors.

While a variety of circuits that may be used for thyratron motor control have been presented and many more are possible, choice of a specific circuit for a particular application will be simplified by the small number of suitable circuits that will remain after available motors, power supply sources, power transformers, and error signals are considered.

Experience shows that it is nearly always profitable to experiment with the filtering to the motor. With bidirectional push-pull circuits it may also be desirable to return the thyratron grid circuit to a tap across the motor armature to obtain a small amount of negative feedback. Such feedback, particularly if through an *RC* circuit, helps to make the grid voltage slightly more negative than normal should all four tubes fire for any reason while the motor is at rest, without seriously affecting the control characteristics; it accordingly helps to reduce power dissipation in the motor for low values of error signal.

12·19. Vacuum-tube Control of D-c Motors.—Hard tubes can be used for direct control of nearly all the types of instrument servo motors for which thyratron control is possible. Circuit details and efficiencies are different for vacuum tubes, however, and it will be seen that they are best suited to certain motor types and operating requirements whereas, for other requirements, thyratrons are preferred. Several characteristics of vacuum tubes are of direct interest in this connection. Among these are the relatively high internal plate circuit voltage drop and, for triodes, the amount by which this tube voltage drop changes with the current through the tube when bias conditions are held constant. For control purposes the triode connection is generally used to minimize static plate resistance and to avoid the constant-current characteristics (high dynamic resistance) of the tetrode or pentode connection.[1] Examples of tubes used for d-c motor control are the triode-connected 6L6 tetrode, which can be driven for short periods to 100 ma at a 100-volt internal drop if the grid is driven $7\frac{1}{2}$ volts positive, giving a drop under this condition equivalent to 1000 ohms in series with the load, and the 2A3-G and 6B4 which will have about the same voltage drop at zero bias. Unfortunately, these are filamentary types; and although suitable for stationary applications, are unsatisfactory for mobile, portable, or airborne equipment. In addition, separate filament transformers may be required for each tube in some circuits, whereas heater-type tubes can withstand 100 volts or more heater-to-cathode potential difference and hence can use a common heater supply.

One interesting tube type is the 6Y6G which has an unusually heavy cathode and will pass 110 ma when triode-connected with an internal

[1] Pentode drive circuits are sometimes used to reduce the time constants of high inductance loads.

drop of 100 volts with the grid about 9 volts negative. Some grid current flows even then, but the resulting loading is not serious. The maximum plate and screen voltage ratings are rather low, 220 and 150 volts, respectively, and may prevent use of the tube where high supply voltages are necessary because of load impedance or generated back emf in cathode-lead circuits.

A new tube type, the 6AR6, made by Western Electric, Tung-Sol, and Hytron, combines the high cathode current of the 6Y6G with the higher anode and screen voltage ratings of the 6L6 and offers good possibilities for improving both the efficiency and power-handling abilities of small hard-tube servo-motor controls.

A second new tube, the RCA-6AS7-G, offers still greater possibilities. This tube, shown in Fig. 15·28, is a low-μ power twin-triode designed particularly for use as a series regulator tube in power supplies. It has the following continuous maximum ratings for each triode unit: plate voltage, 250 volts; plate current, 125 ma; plate dissipation, 13 watts; heater-cathode voltage of either polarity, 300 volts; amplification factor, 2.1; plate resistance, 280 ohms; transconductance, 7500 μmhos; and heater drain, 2.5 amp at 6.3 volts. The plate drop for 100 mils at zero bias is only 30 volts. The glass envelope is the ST-16 size, with a medium shell octal pin base, the same size as a 6L6-G. Availability of this comparatively new tube may be expected to increase greatly the use of hard-tube control of small d-c motors.

A well-known characteristic of vacuum-tube control which should not be forgotten when designing this type of circuit is the grid current that flows in power tubes such as the 6L6 when the grids are driven positive enough to allow the tubes to pass the heavy current required for motor control. A second characteristic, of less practical importance, is the greater control-voltage range that must be used to operate vacuum tubes, as compared with thyratrons, if the entire range from maximum rated power to cutoff is to be covered. For a tube such as the 6L6, for example, a 40- to 50-volt change in grid bias may be necessary to cover the full current to cutoff range, whereas with a 2050 type thyratron, a change in grid bias of roughly 25 volts is sufficient, the exact value being determined by the amplitude of the a-c out-of-phase bias. It is not important, however, that a vacuum tube be completely cut off if used in a bidirectional control circuit where the current flowing in another tube may counteract the effects of a tube not being completely cut off. The statements of Sec. 12·16 with regard to the use of rectified alternating current for d-c motor drive apply directly to vacuum-tube control.

Practical Circuits for Vacuum-tube Motor Control.—A motor that is relatively easy to control with vacuum tubes is the split-series unit. Figure 12·35*a* shows a basic circuit for vacuum-tube control of a split-series motor from an a-c supply using half-wave rectification. Figure 12·35*b* shows the same basic circuit changed for full-wave operation. It requires less filtering than does the half-wave circuit. If a-c control signals are used, they must be fed through transformers or converted to direct current in a previous stage. The full-wave circuit permits higher mechanical power output for the same motor temperature rise.

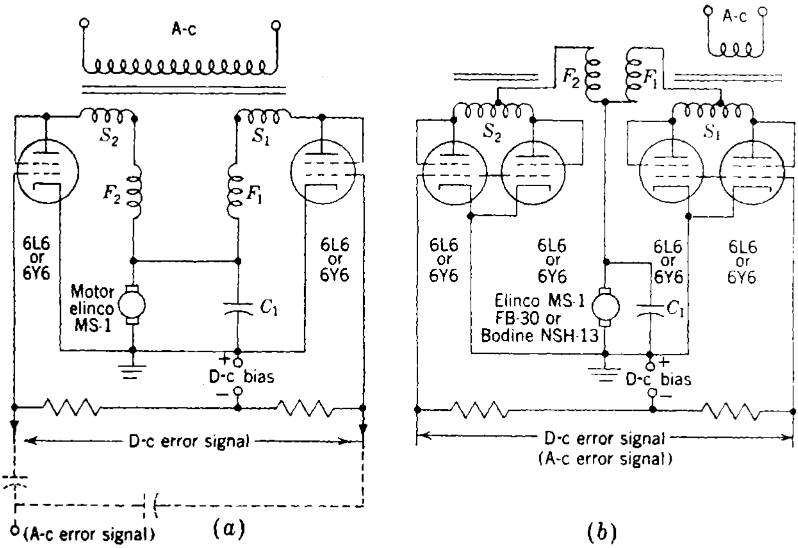


FIG. 12·35.—Control of split-series d-c motor by vacuum tubes. (*a*) Half-wave drive; (*b*) full-wave drive. Either d-c or a-c error signals may be used.

Figure 12·36 gives circuit details for a velocity servo that uses push-pull 6L6's for the motor control. The Elinco B-50 motor is conservatively rated at 0.003 hp when operated at 115 volts direct current. As used in this circuit, a peak mechanical output of the order of 0.005 hp is obtained. It is of interest here, however, to consider the operation of the motor-control circuit. The high-gain amplifier and the tachometer and potentiometer input and output data devices provide a continuous d-c error signal to the drive circuit to bring the motor nearly to the desired speed rapidly and then provide drive to the 6L6's in the form of pulses corresponding to the tachometer ripple. The amplifier is essentially saturated, and the drive circuit is turned full on at the peaks of these tachometer ripple pulses. The fraction of the time that the drive circuit is tuned full on depends on the voltage difference between the tachometer-generator output voltage and the speed control or input potentiometer voltage. This provides effectively a proportional control circuit with the average motor power varying smoothly to give the desired motor speed. A note of caution must be interjected here. In some other circuits, recovery from saturation on tachometer ripple may be too slow, with the result that the amplifier is completely blocked and hence becomes very ineffective. Therefore this must be used carefully.

It is possible to use armature control with hard tubes instead of thyratrons in the circuit of Fig. 12·27. The inherent speed regulation

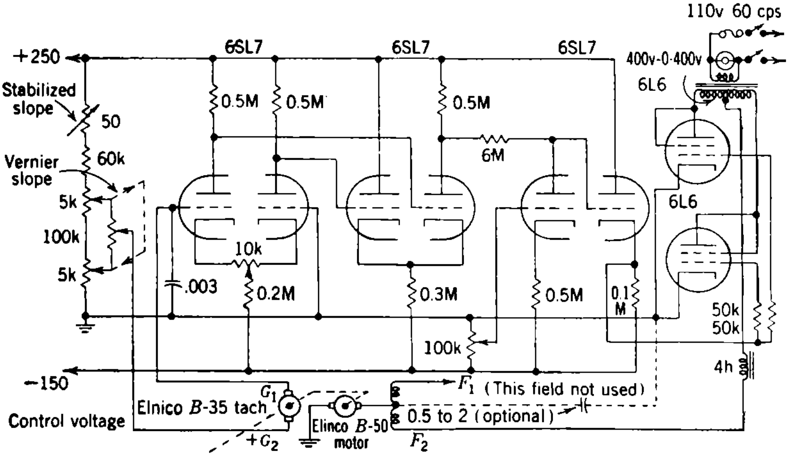


FIG. 12·36.—Full-wave unidirectional velocity servo using 6L6's.

obtained is poorer for the vacuum-tube circuit because the voltage impressed on the armature during the driving half cycle is averaged with the motor back emf during the nonconducting half cycle and fed back into the grid circuit. Since the driving voltage feedback is higher for heavy loads and tends to reduce the drive, larger error signals are necessary to get maximum power from the motor, and top speed drops more with increasing load than for a circuit using thyratrons in which the thyratron firing point is affected only by the armature voltage just before current flows into the armature. Special a-c bias is not necessary for smooth control.

Figure 12·37 gives circuit details of a vacuum-tube motor-control circuit for a velocity servo used in a laboratory test instrument. A complex limit-switching circuit allows push-button control of on-off and reversal. This circuit controls a large (for instrument use) shunt-type motor, the Bodine NSH-12, with push-pull parallel drive to the armature and reverses the direction of motor rotation by reversing both the motor armature and the tachometer-generator output. The motor field is excited from the high voltage d-c amplifier plate supply. Parasitic suppressors such as 25- to 50-ohm resistors or small chokes shunted by

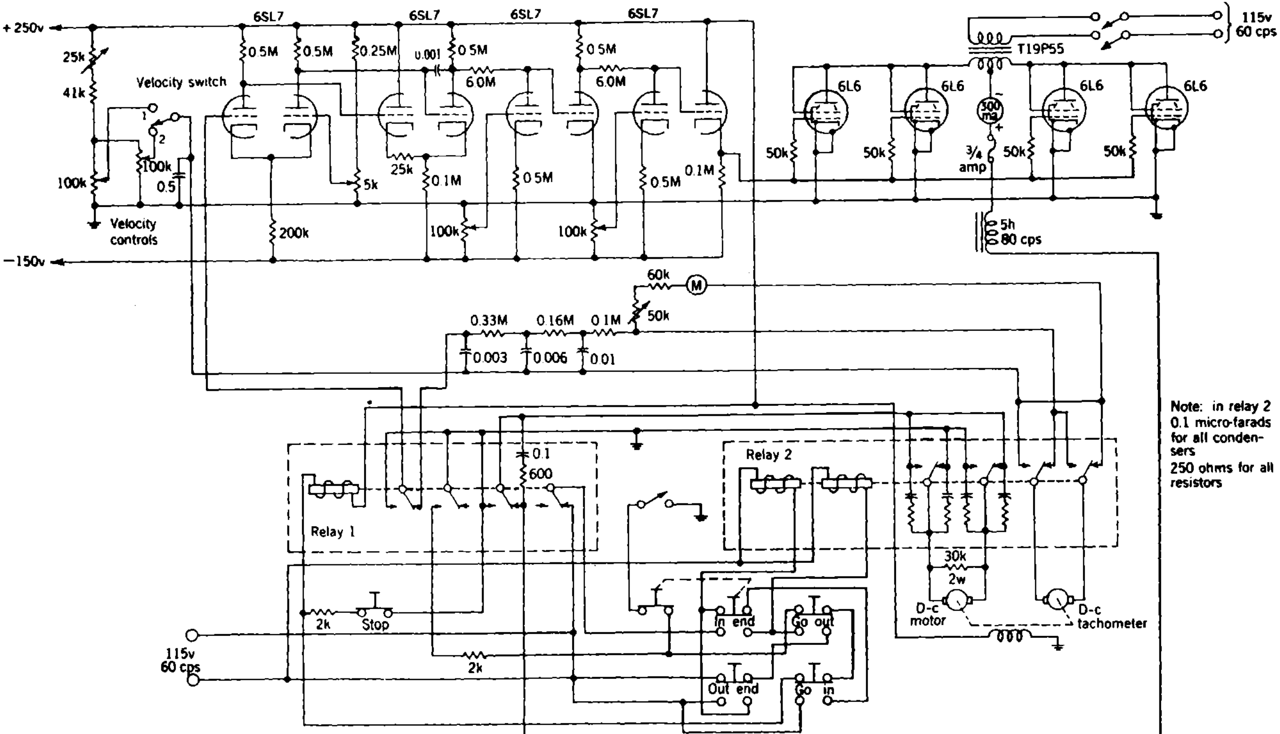


FIG. 12·37.—Range-tracking calibrator servo.

resistors may have to be inserted in the 6L6 screen leads to prevent oscillations. In other respects the control circuit offers no unusual difficulties.

One point may be mentioned here which bears on the quality of speed regulation that can be obtained from a circuit of this type. A small amount of commutation and slot ripple will appear in the generator

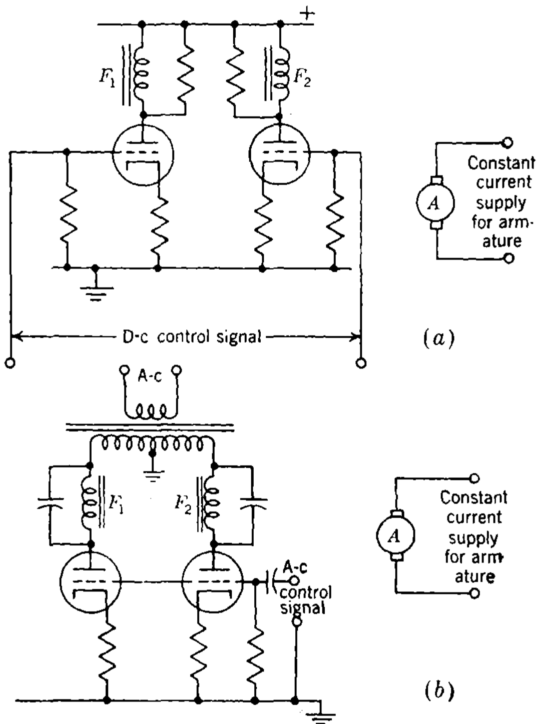


FIG. 12·38.—Basic circuits for bidirectional field-control of d-c motors with vacuum tubes.

output. If the generator is incorrectly "phased" with respect to the motor, its ripple voltage may combine with the variations in motor torque because of motor commutation and armature slot lock to cause periodic variations in speed. The effect is greatest at low speeds and is particularly bad if it is desired to operate below 25 to 50 rpm. A reduction of this effect by a factor of 5 or more may often be obtained by rotating the tachometer generator relative to the motor a few degrees at a time until the smoothest operation is secured.

A difficulty that may be hard to trace is erratic operation caused by a slipping coupling between motor and generator. This usually causes fluctuations in speed or failure of the speed-control circuit to function at all, with the motor running away if unloaded. Reversal of the tachometer leads will also, of course, cause loss of control of motor speed, but this is an easy error to locate and correct.

Another important circuit for hard-tube control of d-c motors uses field control of a motor having a separate field winding for each direction of rotation. Figure 12·38 shows the basic circuit. Field-controlled motors are normally characterized by increase of speed as field excitation drops and by reduced torque at high speed. The most practical compensation for these effects is control of the armature current by a ballast tube or other constant-current device. Saturable-reactor control of a-c power to an armature supply rectifier has been used successfully to secure constant armature current.[1]

For many purposes approximate control of armature current by use of a high-voltage supply and a fixed series resistance large with respect to the armature resistance will suffice, giving a grade of current regulation intermediate between that of a series ballast tube and no compensation. This arrangement, of course, is inefficient because of the relatively large amount of power that is dissipated in the series resistance. When no compensation is used, the armature must be fed from a very low voltage supply to prevent excessive dissipation in the absence of field excitation. This results in low maximum speed and does not normally give a usable control. Figure 12·39 shows speed-torque curves for a typical field-controlled motor with and without armature current control.

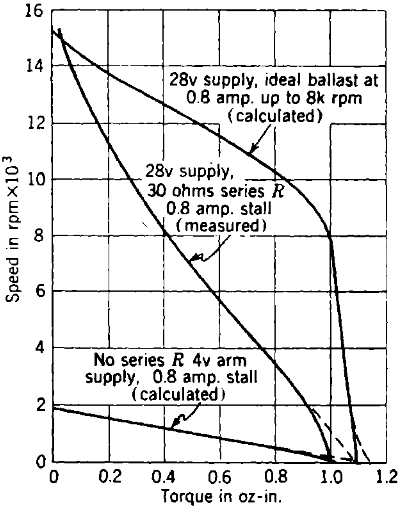


FIG. 12·39.—Torque speed characteristics of Holtzer Cabot RBD-0808 field-controlled d-c motor.

Because of the relative scarcity of motors suitable for field control and because of the attractive features of this method of control, descriptions of several motors together with circuit information, where available, are shown in Table 12·3. A Servomechanisms Laboratory report[2] covers the factors involved in the choice and design of field-type motor control for servo applications in the 75- to 100-watt power output class.

[1] *Fundamental Studies in Servomechanisms*, Vol. II, Part 3, MIT Servomechanisms Laboratory, September 1943, p. 42.

[2] *Ibid.*

An early and successful design of a motor intended specifically for field control is the British Velodyne, which is a split-field motor and tachometer generator mounted on one shaft, designed specifically for velocity-servo applications. As normally used, the motor fields are controlled by the output of a high-gain amplifier. Input to the amplifier consists of a d-c voltage representing the error in instantaneous speed as compared with desired speed or the integral of this quantity. The Velodyne circuit is discussed further in Sec. 12·13.

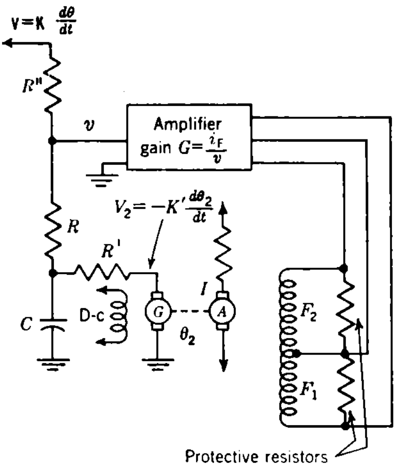


FIG. 12·40.—Block diagram of British Velodyne circuit.

Figure 12·40 illustrates the application of the Velodyne for speed control.

Series armature resistance is used to hold the armature current nearly constant over the range of speed of operation. The control tubes, British type VT-75's, are approximately equivalent to American type 6L6 tubes. Protective resistors of 5000 ohms are shunted across the field to reduce the surge voltage developed when field current is cut off.

Table 12·3 shows the most important characteristics of several sizes of Velodynes and motors without tachometer generators.

Another field-control motor that is particularly useful for instrument servo applications is the Holtzer-Cabot type RBD-0808 unit, recently designed for the Radiation Laboratory for airborne computer use and shown in Fig. 12·17. Normal full-on operating conditions are 6 ma in either 8000-ohm field and 28-volt armature supply with a series ballast resistor to hold armature current at 0.8 amp. Under these conditions a torque output of 0.9 to 1.0 oz.-in. can be maintained up to approximately 8000 rpm. At this speed, the power output is 0.007 hp, comparing favorably with that of the larger d-c and a-c motors. Figure 12·39 shows speed-torque curves obtained for this motor with full field excitation under several armature supply conditions.

Because of the low field current necessary at full power, this motor can be driven with the fields directly in the plate circuits of any small receiving-type triodes such as the 6J6 or 6SN7 types, and consequently a minimum of d-c power is required for the amplifier. The chief drawback of such a motor design in a small unit is the expense of manufacturing such a high-impedance field with suitable insulation and sufficiently well sealed to prevent moisture from penetrating the windings with subsequent corrosion and failure. The motor described has hermetically sealed fields.

TABLE 12·3.—MOTORS FOR FIELD CONTROL*

	Hughes M10 MK II, MK III	Hoover type 86	British Thompson-Houston type Bd-136	Hoover type 73	Scophony Hoover type 74	Hoover type 88
Motor arm., amp	1.55	0.2	4.5	0.2	5	1.25
Field, ma	10	80	100	80	80	45
Tachometer arm., v/rpm	7.5/2000	None	200/4000	60/2000	56/2000	38/6000
Field, ma	PM	None	PM	95	350	16
Stall torque, oz-in	6.9	17	135	17	17	2.1
Power rating, hp at rpm	0.0126/6000	0.08/4000	0.3/4000	0.08/4000	0.08/4000	0.015/6000
Moment of inertia, arm., lb-ft^2	....	....	0.049	0.00356	0.00356	0.0003
Back emf arm, full field, v at rpm	10/3000	....	50/2000	80/2000	....	....
Max. accel., rpm/sec	15000	1200	4400	9000	9000	6000

	Hughes M 15	General Electric, Ltd.-min. servo motor	FHM Ltd.	Holtzer-Cabot RBD0808	Dumore KB†	Oster ES-1-1‡
Motor arm., amp	4.5	4.0	2.5	0.8	....	....
Field, ma	10	10	1000	6	....	....
Tachometer arm., v/rpm	None	None	200/4000	None	None	None
Field, ma	None	None	180	None	None	None
Stall torque, oz-in	....	1.6	48	1.0	....	....
Power rating, hp at rpm	0.015/3000	0.00536/6000	0.2/6000	....	$\approx \frac{1}{4}$	$\approx \frac{1}{4}$
Moment of inertia, arm., lb-ft^2	....	....	....	3.4×10^{-5}	....	....
Back emf arm, full field, v at rpm	....	14/6000	58/3000	....	....	....
Max. accel., rpm/sec	....	10,000	....	45,000	....	....

* Manufacturers: (British) Henry Hughes and Sons, Barkingside, Essex; Hoover, Ltd., Greenford, Middlesex; Scophony, Ltd; British Thompson-Houston, Ltd., Rugby; The General Electric Co., Ltd., Witton, Birmingham; Fractional Horsepower Motors, Ltd.; (U.S.) Holtzer Cabot, Dumore and Oster.

† Fields rewound 3250 turns each of No. 37 wire. (See MIT Servomechanisms Report, *loc cit*).

‡ Fields wound with 4000 turns each of No. 36 wire. (See MIT Servomechanisms Report, *loc cit*).

Induced voltages up to several thousand volts may be developed across the highly inductive control windings of field control motors when the field current is cut off for reversal. To prevent breakdown of the insulation, some shunt protective device should be used to limit the voltage rise. For small motors, it is practical to shunt each field winding with a fixed resistor. This is recommended for the Velodyne motor control where a value of 5000 ohms shunted across the fields had been found satisfactory. Alternatively, this problem may be solved by the use of one or more neon lamps in series, with a striking voltage enough higher than the normal IR drop in the field coil to break down only when excessive voltage is developed. For larger units nonlinear resistors of the thyrite or germanium crystal type may be shunted across the fields. It is also possible to use a diode connected to act as a short circuit for the induced emf. With any of these provisions care must be taken to avoid loading the field in such a way that the time constant for flux decay becomes excessive. A compromise between voltage limitation and flux decay time constant is usually desirable.

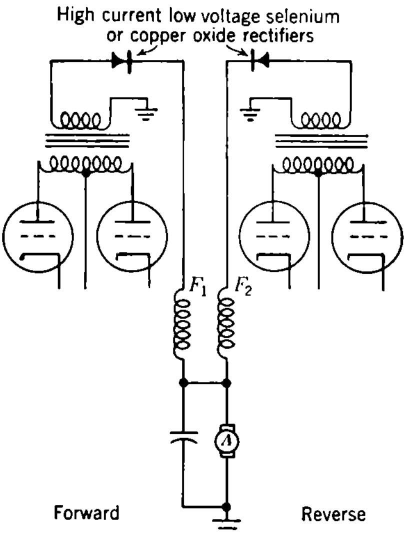


FIG. 12·41.—Transformer and contact rectifier control circuit for split-series d-c motor.

In practice, vacuum-tube control usually has been restricted to conventional split-series motors; until recently there have been few low-power motors suitable for field control or few tubes ideally suited for controlling standard motors. By contrast, suitable split-field motors are relatively cheap and are available from a number of manufacturers in a wide range of sizes. Unless high-impedance motors can be secured, hard-tube control is generally uneconomical. The practical minimum impedance for reasonably efficient direct vacuum-tube control of small motors is probably at about the 110-volt appliance motor level.

Control requirements may permit the use of circuits for unidirectional drive only. For velocity servos, unidirectional drive is often adequate for smooth output, even though the reverse control circuit should be connected for best transient response and perhaps also for extremely low speed operation.

Another possibility for d-c motor control by vacuum tubes is the use of an a-c anode supply for the power-control stage, with transformer coupling of the a-c output to selenium or copper oxide rectifiers supplying a motor. Figure 12·41 is an example of this type of circuit. This basic circuit allows the use of a low-impedance motor with high-impedance vacuum-tube drives; the impedance matching is done by the transformer.

Hard-tube control of a d-c motor from a d-c source as illustrated in Fig. 12·42 may have advantages for some applications, for example, where independence of power-supply frequency is essential. The requirement of rectifiers and filters to provide a sufficiently low impedance source usually limits the use of this type of control to low-power applications or those with very special requirements. If only one d-c bus is available, a split-series motor must be connected in the cathode circuit of the drive tubes for bidirectional drive or the motor must be run with all its terminals above ground potential. If the motor is placed in the cathode circuit of the drive tubes, a high grid-driving voltage is required to overcome the back emf of the armature.

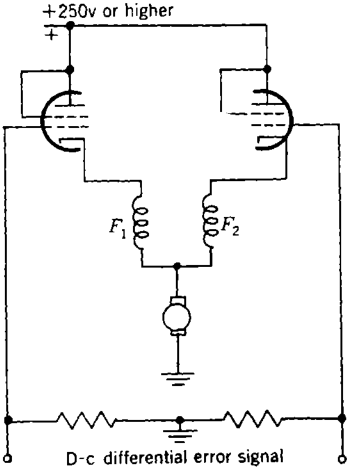


Fig. 12·42.—Circuit for vacuum-tube control of split-series motor from d-c power supply.

12·20. Relay Control of D-c Motors.[1]—This section will consider motor controls in which the power supplied to d-c motors of either the permanent-magnet or wound-field type is controlled by relays. The principal advantage of this method of power control is that the power may be taken directly from a low-impedance d-c source, making unnecessary the heavy and expensive power transformers, rectifiers, filter chokes, etc., that are otherwise needed to supply and control power to a d-c motor. The principal disadvantage is that smoothness and linearity of control are more difficult to achieve with relay control than with other types of control, although not at all impossible, and relay life may be short unless great care is taken. Methods for getting smooth linear and reliable relay control operation will be described. Erratic relay operation is sometimes encountered and has been a serious limitation in a few applications.

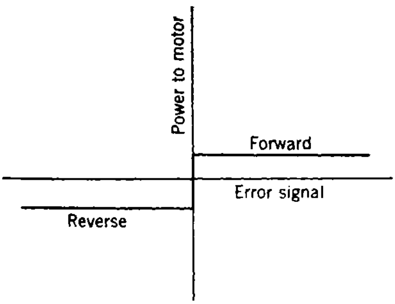


Fig. 12·43.—Relay on-off control characteristics.

Relay servos may be divided into two convenient categories: "on-off" control, in which power to the motor is roughly constant but of polarity corresponding to the polarity of the input error signal to the amplifier, as shown in Fig. 12·43, and "proportional" control in which the power, averaged over a time that is short compared with the mechanical time constant J/f of the motor, is roughly proportional to the input error signal, as shown in Fig. 12·44. Both figures are

[1] Section 12·20 is by W. D. Green, Jr.

very approximate representations of the actual characteristics found in relay servos.

The first type, on-off, is usable only in rather low quality servos, where it is permissible for "hunting" to occur or where the accuracy requirements are such as to allow a dead space large enough to permit the motor to coast to a stop before power is applied to reverse it. Electrically operated brakes have been helpful in making this type of servo useful, by reducing the distance that the motor coasts, after power is turned off.

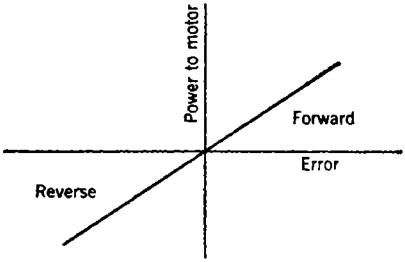


FIG. 12·44.—Relay proportional control characteristic.

The second type, proportional control, when properly designed, can be used in almost any instrument servo application. One way of achieving proportional control with relays is represented by the Westinghouse Silverstat. This is a relay device with a series of moving contacts arranged so that the number of contacts closed is proportional to the current through the control coil. This device can be utilized for proportional control by using the contacts to short-circuit the resistance in series with the motor, using one Silverstat for each direction of rotation.

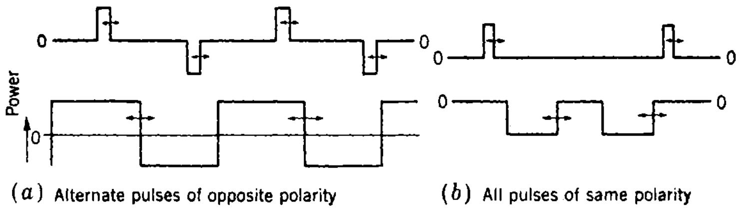


FIG. 12·45.—Typical (idealized) power waveforms for proportional control by variation of pulse length.

Perhaps the best method for proportional control is that in which power is supplied to the motor in a series of pulses of duration short compared with the mechanical time constant to the motor, so that the average power is proportional to the length of the pulse. Then, by making pulse length proportional to error signal input, it is possible to achieve proportional control. There are two possibilities here. One is that power can be applied to the motor first in one direction, then in the other, so that the effective power is the difference between the two (see Fig. 12·45). This method wastes power and heats the motor, so it is usable only where the power involved is small or the efficiency of the motor and control is unimportant or where a suitable filter may be used.

The second possibility is that the two pulse lengths be nearly zero for zero power, with the proper pulse increasing in width depending on the amount and polarity of net power that it is desired to apply to the motor. This action can be achieved in the following way: Consider a pair of single-pole, double-throw relays, connected as shown in Fig. 12·46. It can be seen that if both moving contacts C_1 and C_2 are against fixed contacts a_1 and a_2 or against fixed contacts b_1 and b_2, respectively, there is no voltage applied to the load. If, however, C_1 is against b_1 and C_2 against a_2, the voltage E is applied to the load. Similarly, if C_1 is against a_1 and C_2 against b_2, the voltage E is again applied to the load, but in the opposite direction. Now consider what happens when C_1 and C_2 are caused to move back and forth between a_1 and b_1 and a_2 and b_2, respectively. If both move from a to b at the same time, no voltage is applied to the load; but if C_1 precedes C_2, voltage is applied for the period that C_1 is at a_1 and C_2 is at b_2. This is shown graphically in Fig. 12·47.

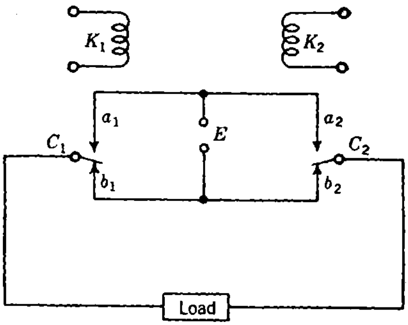


FIG. 12·46.—Contact connections for bidirectional control of d-c motor with two relays.

As a means of obtaining the action described above there may be used two currents $i_1 = i_{dc_1} + I \sin 2\pi ft$ and $i_2 = i_{dc_2} - I \sin 2\pi ft$ flowing in the coils of the two relays, respectively, where i_{dc_1} and i_{dc^2} are fixed currents and I is the amplitude of the sinusoidai current flowing in both relays. This is illustrated in Fig. 12·48. The circuit is shown in Fig. 12·46. The contacts C will move back and forth between a and b. If i_{dc_1} is increased and i_{dc_2} is decreased, C_1 will move to a_1 before C_2 moves to a_2. Furthermore, C_1 will remain at a_1 longer than C_2 will remain at a_2. The result is a series of pulses, at twice the flutter frequency, of duration roughly proportional to the difference between i_{dc_1} and i_{dc_2}. The exact relation is a function of i_{dc_1}, i_{dc_2}, I, i_{op}, i_{re}, f, and the resonant properties of the relays. A simple analysis, neglecting the mechanical characteristics of the relay and assuming that the control currents in the two relays

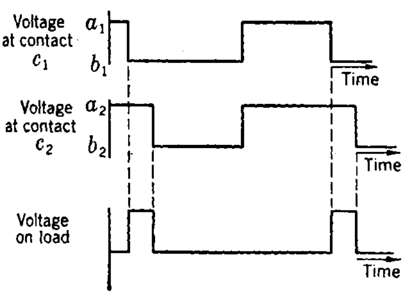


FIG. 12·47.—Voltage waveforms at relay contacts in control circuit for bidirectional control of d-c motors.

are symmetrical with respect to the average of the operating and release currents ($i_{dc_1} + i_{dc_2} = i_{op} + i_{re}$), leads to a characteristic somewhat like that shown in Fig. 12·49*a*.

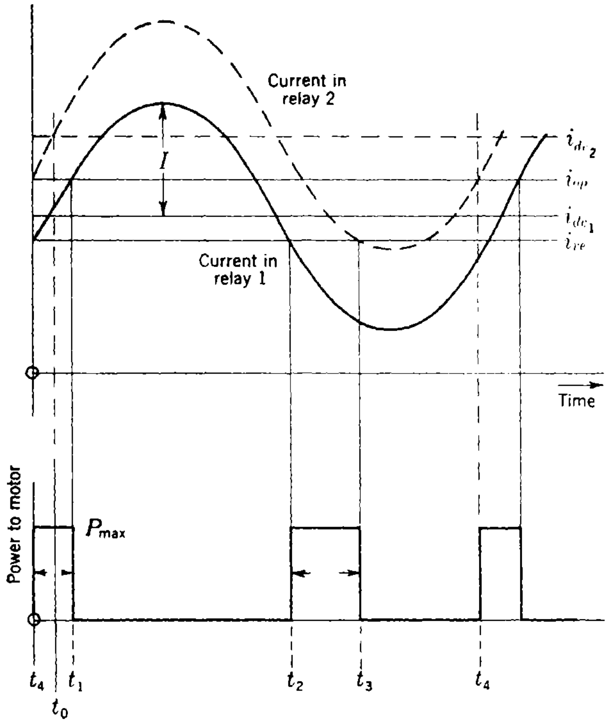


FIG. 12·48.—Graphical analysis of proportional control characteristics.

If i_{dc_2} and i_{dc_1} were symmetrical about the average of i_{op} and i_{re}, the resulting pulses of power to the motor would be of equal width.

$$i_{re} - i_{dc_1} = I \sin \frac{2\pi t_2}{T},$$

$$i_{re} - i_{dc_2} = I \sin \frac{2\pi t_3}{T},$$

$$t_2 = \frac{T}{2\pi} \sin^{-1}\left(\frac{i_{re} - i_{dc_1}}{I}\right),$$

$$t_3 = \frac{T}{2\pi} \sin^{-1}\left(\frac{i_{re} - i_{dc_2}}{I}\right),$$

$$t_3 - t_2 = \frac{T}{2\pi}\left[\sin^{-1}\left(\frac{i_{re} - i_{dc_2}}{I}\right) - \sin^{-1}\left(\frac{i_{re} - i_{dc_1}}{I}\right)\right],$$

$$t_1 - t_4 = \frac{T}{2\pi}\left[\sin^{-1}\left(\frac{i_{op} - i_{dc_1}}{I}\right) - \sin^{-1}\left(\frac{i_{op} - i_{dc_1}}{I}\right)\right].$$

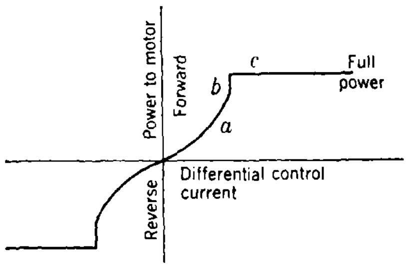
For the symmetrical case (not illustrated),

$$
\begin{aligned}
i_{op} - i_{dc_1} &= -(i_{re} - i_{dc}),\\
i_{op} - i_{dc_2} &= -(i_{re} - i_{dc_1}),\\
t_3 - t_2 &= t_1 - t_4,
\end{aligned}
$$

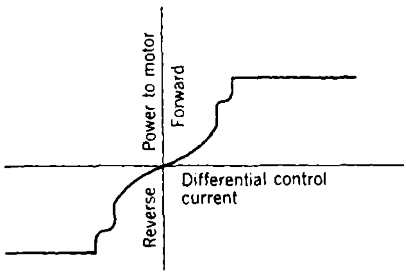
and

$$P_{avg} = \frac{2(t_3 - t_2)}{T} P_{max}.$$

In Fig. 12·49 a plot of power applied to motor vs. differential control current $i_{dc_1} - i_{dc_2}$ is shown. The part of the curve labeled a is a modified



(a) Control currents symmetrical about average of operate and release currents.



(b) Control currents asymmetrical with respect to average of operate and release currents.

FIG. 12·49.—Characteristics of proportional control with relay.

sine function. Part b is an abrupt increase which occurs when $(i_{dc_1} - I)$ becomes greater than i_{re} and $(i_{dc_2} + I)$ become less than i_{op}. If the i_{dc} currents are not symmetrical with respect to the mean of the operating and release currents, there will be two such discontinuities in the characteristic, as shown in (b). Part c of the curve represents full power applied to the motor. The magnitude of part b is the full magnitude of c times $2\cos^{-1}\left(1 - \frac{i_{op} - i_{re}}{I}\right)$. Consequently, I must be large compared with $(i_{op} - i_{re})$ if it is desired to extend the proportional control to its limit.

As emphasized previously, this is a somewhat oversimplified analysis. It serves, however, to outline the general philosophy of this type of control. A note regarding the effect of the frequency f is in order. Because of the inertia of the moving contacts, some time must be spent between fixed contacts, so that, at higher frequencies, a smaller proportion of the time is available for power. Also, the mechanical motion of the contacts in general will lag in phase the driving current in the coil; this effect increases with frequency but is complicated by resonance phenomenon. It should also be pointed out that characteristics of the same model relay may differ appreciably from one unit to the next, which further complicates analysis. It is possible to use waveforms other than sinusoidal to cause the relay contacts to vibrate. Because ordinary relays are relatively slow moving, a special, fast-moving relay is desirable for the type of use described above. Such a relay is the Western Electric D-168479, known also as the type 80. This is a single-pole double-throw relay, whose contacts are enclosed, under pressure, in a sealed-glass cartridge. A sketch of this relay is shown in Fig. 12·50. The contacts are wet with mercury, which ensures large contact area, reduces bouncing of the contacts, and enables the unit to handle large currents. The moving contact, which is fixed at the lower end, touches a pool of mercury, which, by a wetting action, deposits itself in a thin film over all the contacts. The magnetic field that actuates the relay is applied longitudinally. The coil windings are in the form of a hollow cylinder, inside which the glass cartridge is placed. The whole assembly is placed in an octal base metal tube, resembling very closely a metal 6V6. As supplied by Western Electric, these relays have two coil windings, so that the d-c control current can be applied through one and an a-c "buzz" current through the other. However, there is no reason why the two cannot be added electrically, and it is a common practice to use the two coils in series. This relay is capable of operation at frequencies in excess of 100 cps and of controlling instrument servo motors of all sizes. It is necessary to go to some lengths to protect the contacts when large powers are involved. There are two principal methods for making the contacts vibrate. One is that suggested above, in which some external source of alternating current is used to vary the current

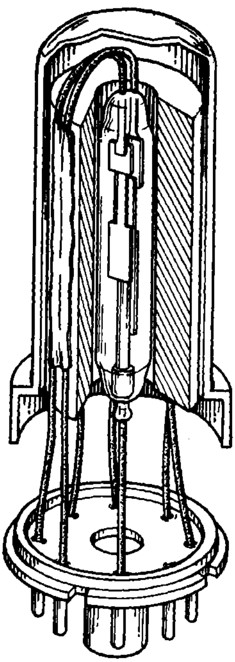


FIG. 12·50.—Western Electric D-168479 mercury contact relay.

in the relay coil. This external source can be a separate oscillator: or where 60 cps is the power frequency, a 6.3 volt filament supply may be used for this purpose. The a-c can be applied directly to the relay coils, or it can be introduced at the grids or cathodes of the amplifier.

A second method, especially useful if a suitable source for a-c is not already present, or if economy of weight, size, or even expense demands that parts be kept at a minimum, is that in which, by suitable regeneration, the relays are made part of a self-oscillatory circuit. This circuit

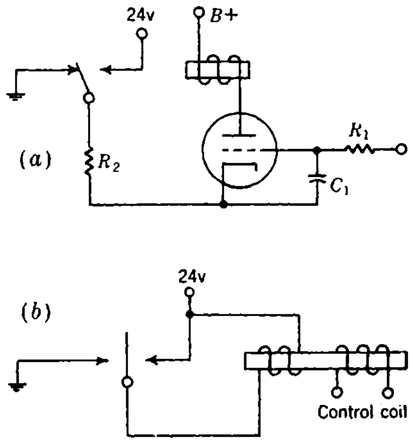


FIG. 12·51.—AN/APA-46 antenna-control servo using two mercury contact relays.

bears considerable resemblance to an ordinary doorbell or buzzer. Again, this can be accomplished through connections either to the coils or to the amplifiers. Figure 12·51 shows circuits for achieving this self-oscillation. It can be seen that if the grid is at a potential such that current through the relay coil is somewhat more than enough to operate the relay, the relay will operate, raising the cathode potential and, through condenser C_1, the grid potential. C_1 will discharge through R_1, and the potential of the grid will drop, reducing the plate current until the relay releases. When this occurs, the cathode potential is reduced, as is the grid, through C_1, but again C_1 charges, and the grid potential rises, increasing the plate current, and the relay again operates. This oscillation continues at a frequency dependent on R_1C_1 and also the mechanical and electromagnetic time constants associated with the relays. This type of circuit was used quite successfully in the AN/APA-46 antenna servo,[1] where this arrangement was used on both halves of the differential amplifier driving the relays. The equipment is shown in

[1] Designed by Radiation Laboratory, produced by Thomas B. Gibb Co., Delavan, Wis.

Fig. 12·52, and the circuit is shown in Fig. 12·53. It is true that the signal changes the oscillation slightly, but this is of no great importance.

With these relays, there is invariably a very short period of bridging between the two fixed contacts as the moving contact changes from one fixed contact to the other. Suitable means must be provided to limit the current that flows under this condition. This action also modifies slightly the operation of the buzz circuit described.

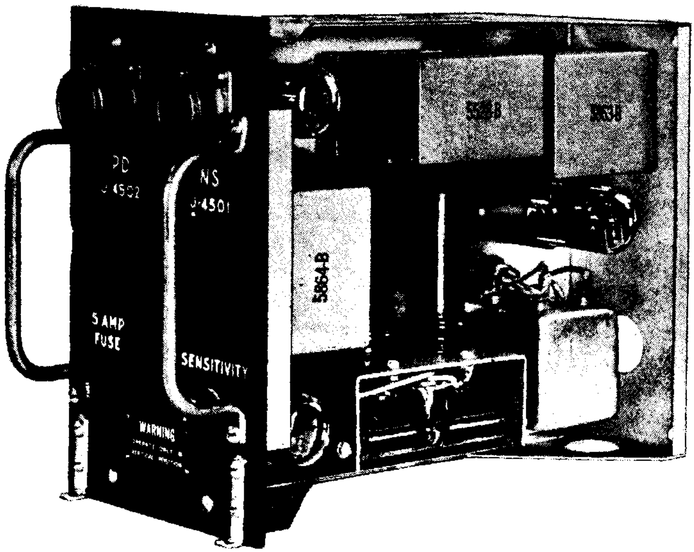


FIG. 12·52.—AN/APA-46 airborne radar antenna-control servo.

In spite of the fact that the response of the relays system described is not so rapid as vacuum-tube or thyratron control, it is much more rapid than the motor in the usual circuit and hence gives smooth and fairly linear control of motor power. The use of relays in the manner described has little effect on the design of the rest of the servo. The worst limitation of relay servos has usually been a tendency toward erratic operation of the relays. The self-oscillation buzz circuits aid somewhat in reducing the extent of erratic operation.

An alternate method of achieving approximately linear control from a relay or other on-off type control device is to allow the servo itself to oscillate at a high frequency but with low amplitude by suitable adjustment of servo loop parameters. MacColl[1] has discussed both the auxil-

[1] L. A. MacColl, *Fundamental Theory of Servomechanisms*, Van Nostrand, New York, 1940, pp. 78–87.

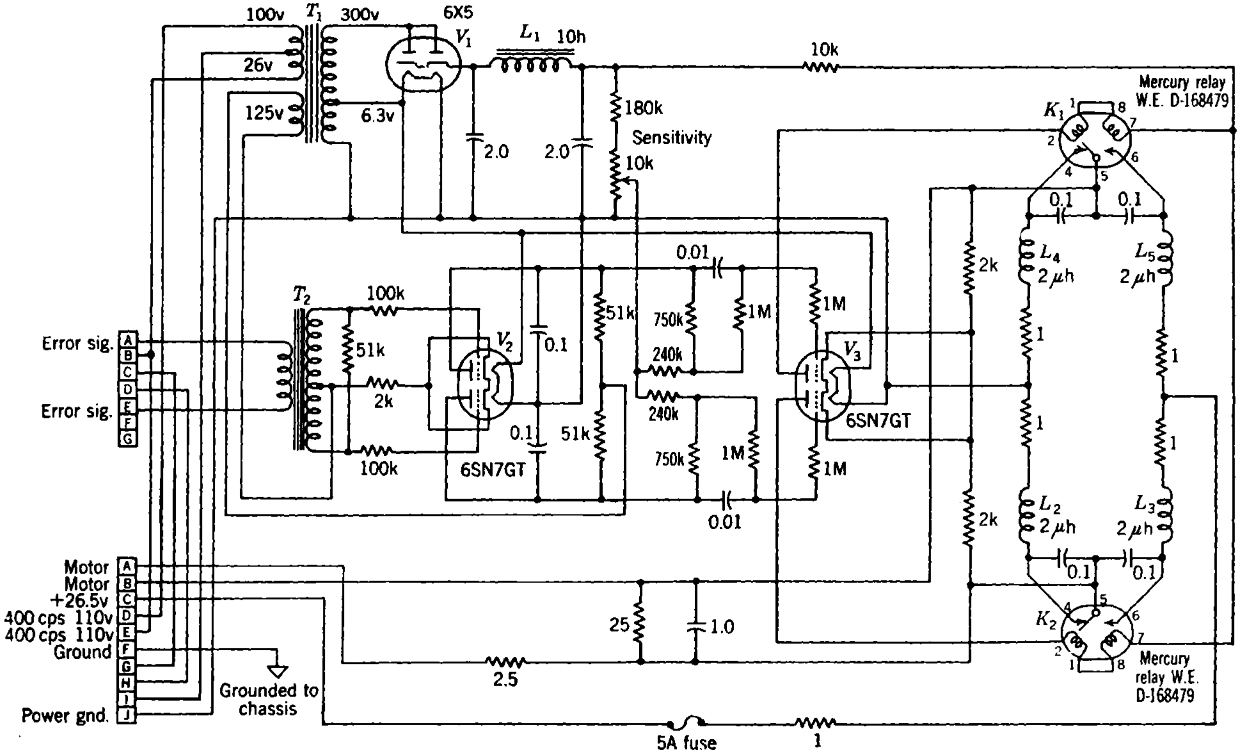


Fig. 12·53.—Schematic of AN/APA-46 antenna-control servo.

iary sinusoid and the oscillating system method of using on-off servos to approximate linear control.

One of the most important considerations in the design of a relay servo is the protection of the contacts. Because of the inductive motor armature and the presence of back emf, this is sometimes a difficult problem. In general, it is necessary, by the use of combinations of capacitors and inductors, to limit the rate of change of current through them, as well as limiting the peak currents. The manufacturers have made specific recommendations for this protection.[1]

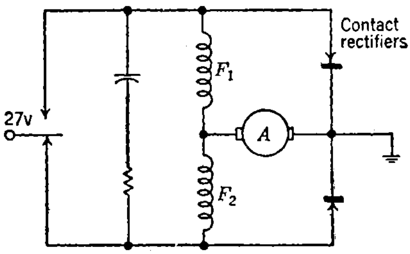


FIG. 12·54.—Relay contact protection with small rectifiers.

A method of contact protection is illustrated in Fig. 12·54. This circuit uses small contact rectifiers to limit the voltage peaks when the current through the motor is interrupted by the relay contact.[2] Other relays require similar protection. In addition, there can arise mechanical troubles, as a result of the hammering of the contacts. This, however, is a question of relay design and is not something that can be remedied by circuit design. It should, however, be borne in mind when selecting relays.

12·21. Controlled Generators.[3]—Motor-speed control can obviously be obtained by varying the voltage of the line supplying power to the motor. Several practical ways of doing this have been proposed, and two or three have achieved rather wide use for applications requiring output powers between 100 watts and 1 hp. The general plan is to use vacuum-tube control of the field current of a d-c generator, using the generator output to operate a d-c motor. A high power gain can be achieved in the controlled generator, and consequently a pair of 6L6's or similar tubes may give adequate drive for ultimately controlling a gun turret or other heavy load. The chief disadvantage of this system, and one that becomes increasingly important for low-power applications, is the need for three rotating machines: a motor to turn the controlled generator, the generator itself, and the final servo motor. For some applications the time lags involved in establishing and in reversing the generator flux are excessive; this is particularly true of the Amplidyne

[1] Bell Telephone Laboratories *Report No.* 741.

[2] This circuit has been used extensively in servos designed by Bell Telephone Laboratories.

[3] Section 12·21 is by J. R. Rogers and I. A. Greenwood, Jr.

generators where a short-circuited armature provides the effective magnetic flux for generation of the usable d-c output.

For these reasons, and particularly because of their size, controlled generators are usually out of the question for instrument servo applications. For completeness, however, the important types will be mentioned briefly, and notes on their range of sizes and peculiar characteristics will be given below.

The *Ward-Leonard System* consists of a vacuum-tube-controlled field on a d-c generator, the output of the d-c generator being used to operate a d-c motor. An a-c or d-c motor is used to turn the control generator. This system suffers from the large amount of equipment required. For applications where this disadvantage is not serious, the system has been used extensively and has given fairly satisfactory performance.

Amplidyne Generator Systems[1] are a trade-marked product of the General Electric Company. A low-power generator control field is energized by the output of a vacuum-tube amplifier. A nearly conventional generator armature is used, but there are two sets of brushes 90° apart. The brushes in the position for generating output from the control field flux are short-circuited. A reaction flux is set up at 90° from the control field flux, and this acts on the armature to give a generator output voltage from the second pair of brushes, the output being in proportion to the current flowing in the control field. The Amplidyne differs from the Ward-Leonard system in that a high gain is obtained by the use of the short-circuited brush connection in the controlled generator so that larger loads may be controlled with smaller vacuum tubes. These units are available in capacities from 75 watts to 5-kw output. The generators and drive motors for the small sizes are available mounted as a single motor-generator unit. A variety of circuit arrangements have been worked out for stabilizing Amplidyne servos; in general these involve feedback from the generator output or from the output load current to compensating windings on the control field poles.

12·22. Vacuum-tube Control of A-c Motors.[2]—Perhaps the most common method of controlling a-c motors for servo applications is by the use of vacuum tubes. This method is especially applicable to the control of small two-phase squirrel-cage induction motors such as the Diehl FPE 25-9 but is also used in the control of other types of motors.

Since all the power controlled by a vacuum-tube power-control circuit must pass through the output tubes, the method is necessarily inefficient and bulky. It is to be noted, however, that half or more of the power of a two-phase motor comes directly from the line and need

[1] For details see, for instance, J. R. Williams, "Amplidyne Characteristics and Operating Principles," *Elec. Eng.*, **65**, 208–13, May, 1946.

[2] Section 12·22 is by W. Goodell., Jr, and I. A. Greenwood, Jr.

not be controlled. Large tubes and heavy transformers must still be used if the load is appreciable. With a 60-cycle power frequency the transformers and filter chokes and condensers needed are particularly bulky and heavy. Since considerable power is dissipated in the tubes, the problem of cooling them must be considered. The most important consideration in the design of a vacuum-tube power-control circuit for the control of an a-c motor is its output impedance. If the motor-control circuit is designed for a particular motor, the output impedance should be matched to that motor under normal load conditions. If the amplifier is to be used with any of several unspecified motors, the output impedance is generally made as low as is convenient. Because of the relatively low impedance of most a-c motors, it is usually necessary to use an impedance-matching transformer in order to obtain the desired amount of power in the load with the vacuum tubes that are available. Very low impedance tubes, such as the 6AS7G (r_p of 280 ohms at $e_c = 0$), may sometimes be used without transformers.

A second important consideration in the design of power-control circuits for a-c motors, and in particular for control of two-phase induction motors, is the question of phase shift. Where frequency is sufficiently constant, it is convenient to tune the load—the load in this case being the control winding of the motor as seen through the impedance-matching transformer. The tuning may be done by connecting a condenser either across the motor winding or on the primary side of the motor-driving transformer; the advantage of the latter connection is that a smaller size capacitance is required. The tuning also attenuates harmonics of the carrier frequency.

It is appropriate at this point to warn the reader of the complexities involved in a complete understanding of the phase and amplitude relations in the two windings of the motor as speed, torque, and control voltage are varied. A particularly troublesome point is the change of power factor and impedance due to back emf. In practice, it is usual to determine the optimum shunt capacitance empirically. Under some conditions, the optimum capacitance will differ considerably from that which tunes the motor-winding inductance for the a-c frequency used. Maximum torque is produced when the currents in the two windings differ in phase by nearly 90°. Factors that must be considered are the complex characteristics of the motor itself, the source impedance, the shunt or series capacitance, the load conditions that will be imposed on the motor, and sometimes nonlinearity in the operation of the control element or motor. Most of these difficulties may be resolved by simple experimental study, and those remaining are minimized in importance by the fact that the control circuit and motor are part of a feedback loop in which high gain and accurate linear feedback minimize irregularities in the μ-portion of the feedback loop.

In circuits where the alternating current applied to the controlled winding of the motor is obtained from amplification of the error signal, the necessary phase-shift network is occasionally introduced in the error voltage circuits prior to the power-control stage.

There are many methods by which a-c motors may be controlled by vacuum tubes, a few of which are listed and described below. Figure 12·55 shows the circuit for a general-purpose amplifier and a general-

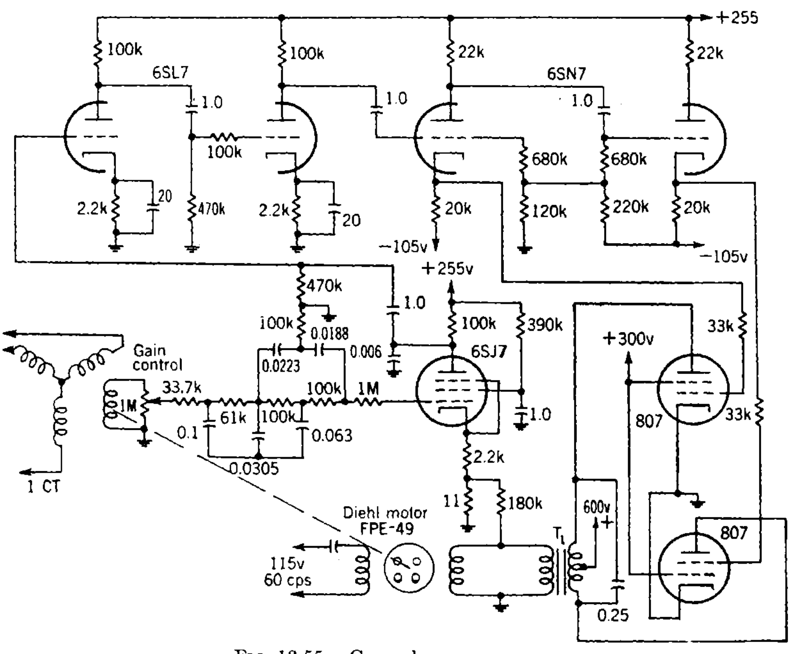


FIG. 12·55.—General-purpose servo.

purpose all a-c servo circuit, capable of driving a variety of motors and of working smoothly with a wide range of loads and speeds. It consists of a 60-cycle carrier stabilization network, the constants of which may be changed to fit the particular load used; a three-stage amplifier with a gain of 200,000 (106 db); a phase inverter; two push-pull 807 output tubes; and an output transformer. Negative feedback is employed across the transformer and amplifier to cut the gain to 12,000 (82 db) and to reduce the output impedance to about 33 ohms. Any 60-cycle two-phase induction motor with a voltage rating for the control phase of less than 150 volts and of less than 50 watts rated output may be used, although the specific motor for which the stage was designed was the Diehl FPE-49 low-inertia two-phase squirrel-cage induction motor.

Figure 12·56 shows the output stage of a 400-cycle airborne servo driving a Diehl FPE-25-9 induction motor. The plates of the push-pull output tubes are connected through the output transformer to the control winding of the motor. The grids of the output tubes are fed with push-pull alternating voltage proportional to the error signal. The load is tuned by a condenser across the primary of the transformer.

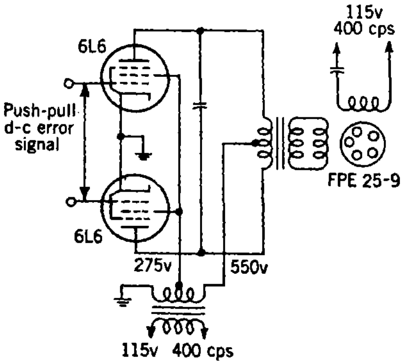


FIG. 12·56.—Airborne servo power stage for 400-cps operation of induction motor.

Figure 12·57 shows the output stage of a small 400-cycle servo. The two output tubes act as modulators for the output current, the phase of the output voltage depending on the phase of the a-c input error signal. The condenser across the motor field serves to smooth out the a-c wave shape as well as to tune the motor.

Figure 12·58 shows a typical single-ended output stage used in a ship-borne 60-cycle system. Feedback around the transformer is used to provide additional stability and linearity of control.

Figure 12·59 shows how vacuum-tube control may be applied to field-controlled universal-type motors. The output tubes are connected

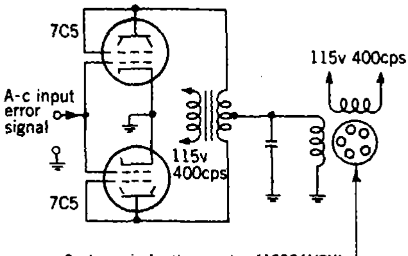


FIG. 12·57.—Output stage of Brown Instrument Co. 400-cps servo.

differentially to the split field of the a-c motor. Brush noise is apt to be troublesome in this circuit. Brush life should be carefully checked, as wear may be rapid.

12·23. Alternating-current Control with Saturable Reactors.[1]—The major disadvantages associated with vacuum-tube control of a-c motors are inefficiency and rather great space and weight requirements. These disadvantages result from the fact that in this type of control all the

[1] Section 12·23 is by S. B. Cohen and I. A. Greenwood, Jr.

power delivered to the controlling winding of the motor must pass through a vacuum tube or tubes. Elimination of this condition is a feature common to saturable reactor, saturable transformer, and relay control of a-c motors. In these types of control small amounts of power may be

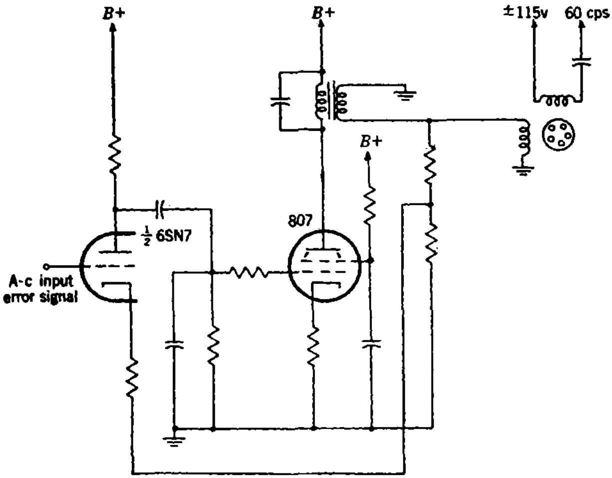


Fig. 12·58.—Single-ended motor power-control stage.

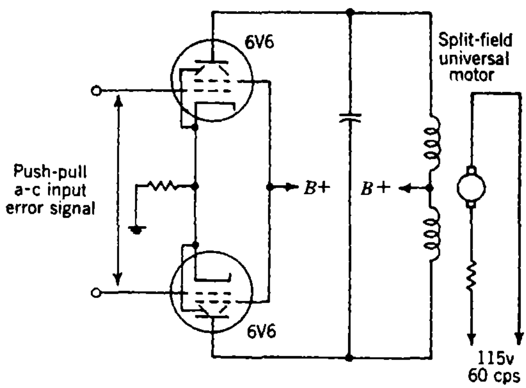


Fig. 12·59.—Vacuum-tube control of split-field universal-type motor operated on a-c.

used to control the large amounts of power necessary to operate motors. Relay control of a-c motors is discussed in Sec. 12·24. Saturable transformers followed by contact rectifiers are sometimes used for the control of d-c motors.

Both saturable reactors and saturable transformers are used for controlling a-c motors. Figure 12·60 shows in a simplified form one type of saturable reactor. This unit has a core with three legs; the outside two have a-c windings which are wound either in series aiding or in parallel opposing, while the center leg has a d-c or control winding. Sometimes an additional winding is also put on the center leg to produce a flux bias to set the operating point of the *B-H* curve. The flux of the two a-c windings cancels in the center leg. As the d-c control current increases, the two a-c legs will saturate. This saturation causes the reactance of the a-c windings to be reduced. Thus, when the d-c control current is varied, the reactance as seen at the a-c terminals will vary over an

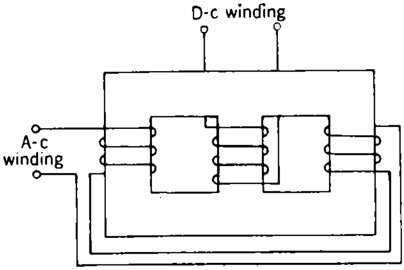


FIG. 12·60.—Simple saturable reactor.

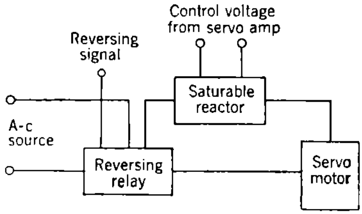


FIG. 12·61.—Use of saturable reactor in series with load.

appreciable range.[1] The ratio of maximum to minimum reactance obtainable with this type of saturable reactor is about 25/1, although larger ratios have been obtained with very special core materials and coil arrangements. The ratio of the d-c power used in a saturable reactor to the power controlled by it may run from 1/15 to 1/35.

Figure 12·61 is a block diagram showing an application of one of these reactors in a servo system. The reactor is in series with the motor load. As the d-c control voltage from the servoamplifier is varied, the reactance of the saturable reactor varies and thus the voltage across the motor varies. The motor may be a two-phase induction motor with one winding excited by a voltage 90° out of phase with the controlled phase, or other a-c motors may also be used. If bidirectional operation is desired with this type of control, a relay and sense-detecting circuit must be used to make the motor operate in the proper direction. With these limitations the circuit as shown would be appropriate for velocity control but not for position control.

Since the reactance varies smoothly over its range, smooth but not necessarily linear control may be obtained with this type of power-control element. The time constant introduced by the reactor sometimes limits the speed of operation of this type of system. A serious limitation

[1] *Fundamental Studies in Servomechanisms*, MIT Servomechanisms Laboratory, Vol. II, Sec. 3.

in the application of linear servo theory is encountered in connection with saturable reactors, for they are usually quite nonlinear.

The disadvantages of using a relay to reverse the motor rotation may be eliminated by using a type of control shown in Fig. 12·62. This system consists of a conventional servoamplifier which drives two separate saturable transformers so connected that at balance the output voltage to the motor is zero. In operation, the output voltage will reverse its phase relative to the line voltage as one transformer becomes more saturated than the other. A saturable reactor servomechanism used extensively in airborne application is shown in Fig. 12·63. These units are produced under the trade name of Torque Unit and Torque

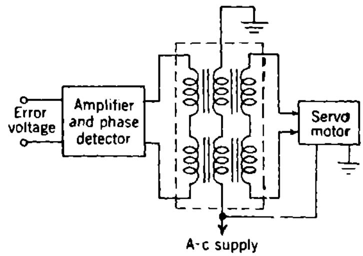


FIG. 12·62.—Use of saturable transformer in servo system.

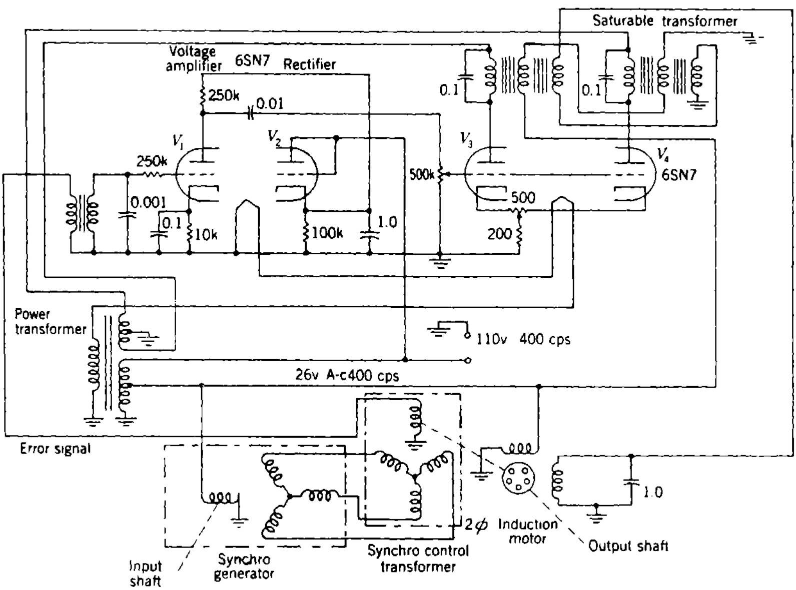


FIG. 12·63.—A 400-cps servo using Bendix saturable transformer.

Amplifier[1] and were used in the AN/APS-15 and AN/APQ-13 radars, in the Fluxgate Compass system, and in other airborne applications.

The saturable transformer unit, in the plate circuits of tubes V_3 and

[1] Trade names of the Eclipse-Pioneer Division of the Bendix Aviation Corp., Teterboro, N.J.

V_4, consists of two transformers with separate cores connected so that under balance conditions no voltage is applied to the controlled winding of the induction motor. Any error signal present is amplified by tube V_1 and further amplified and phase-detected by tubes V_3 and V_4. The resultant direct current flowing in the saturable transformers tends to increase the voltage output of one transformer and decrease the voltage of the other transformer, resulting in voltage being applied to the controlled winding of the induction motor. The parallel capacitor shifts

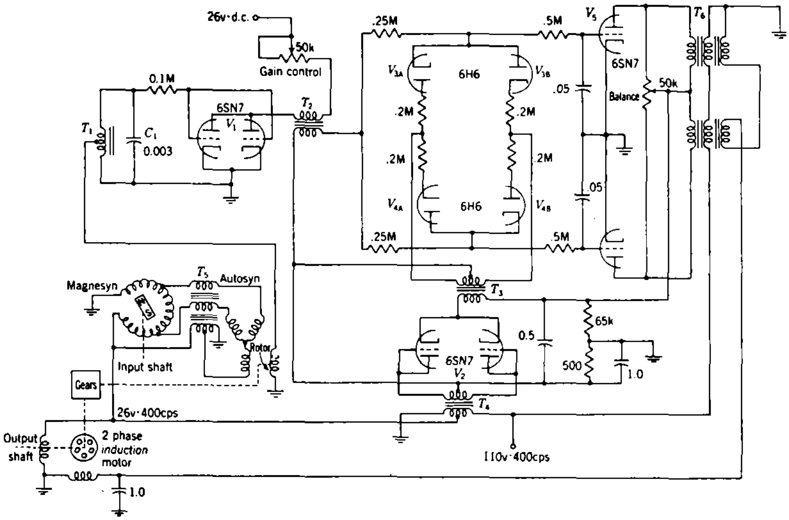


FIG. 12·64.—Bendix "Torque Amplifier" using saturable transformer and Magnesyn.

the phase of the current in the controlled winding. The saturable transformer unit, consisting of the two separate transformers mounted together, weighs approximately 18 oz and may be used to control a Bendix CK-1 or CK-5 motor. For stabilization purposes, the motor usually includes a small magnetic damper, the equivalent of viscous damping on the output shaft. For highest performance, stabilization may be achieved by an induction-type a-c tachometer used with its output signal added to the error signal by series connection.

A number of variations of the servo shown in Fig. 12·63 are available, including that shown in Fig. 12·64 which is a servo system using a Magnesyn as the input data device and an Autosyn as the data output device.

Transformer T_5 removes signals at the exciting frequency from the Autosyn. A double-frequency signal for reference purposes is derived from tube V_2, applied to the diode bridge formed by tubes V_3 and V_4, to which the amplified signal from tube V_1 is also applied. The resulting direct current is amplified in tubes V_{5A} and V_{5B} and controls the relative saturation of the two transformers T_6, thereby controlling the rotation of the motor.

12·24. Relay Control of A-c Motors.[1]—Relays offer additional methods for the control of a-c motors. In general, their use is much the same as that described in Sec. 12·20 for the control of d-c motors. Relay control has the advantages of less power dissipation in the control unit, less weight, and smaller size as compared with vacuum-tube, thyratron, or saturable reactor control of a-c motors. On the other hand, as before, smoothness and accuracy are less than are obtainable with these other methods. Nevertheless, there are a great many servo applications where the requirements can be more than adequately met by relay control.

The problem of proportional control in a-c servos is somewhat complicated by the relation between the power frequency and the frequency of vibration of the relay contacts. Unless the difference between these two frequencies is large compared with the frequency at which the motor can respond, erratic motion may result. Consequently, if the power frequency is 60 cps, the contact frequency would have to be quite low, say 10 or 15 cps, or considerably higher than 60 cps, say 100 cps. Most commercially available relays are incapable of the higher frequency, an exception being the Western Electric D-168479 relay discussed in Sec. 12·20. Another possibility, however, is to have the contact frequency the same as the power frequency. As before, the proportionality is achieved by variation of the length of the power pulse.

Except for its effect on the frequency of contact vibration, the use of a-c power has little effect on the design of the amplifier and relay control circuits. Protection of the relay contacts is easier than with direct current, since alternating current does not sustain an arc. As a further refinement of contact protection, it is even possible to do the switching when the instantaneous current through the contacts is zero.

Figure 12·65 illustrates a control unit comprising two 2D21 miniature thyratrons and two Western Electric D-168479 mercury contact relays of the type discussed in Sec. 12·20. This control unit is designed for a servo application that does not require highly accurate positioning or particularly fast response to input error signals. The entire unit is extremely compact and can be built to fit into a 4-in. cubical space. It can be used to drive a large variety of a-c motors, although the particular motor for which it was designed is a Diehl FPE-49 two-phase 60-cps squirrel-cage induction motor, which has a rating of 11 watts mechanical output.

The plates of the thyratrons are energized from the center-tapped secondary of a step-up transformer, and the grids are energized by the

[1] Section 12·24 is by W. Roth and W. D. Green, Jr.

60-cycle error voltage. The voltages on the respective plates are 180° out of phase while the grid signals are in phase; therefore, one thyratron will conduct during the time that its plate voltage is positive for one polarity of input signal, while the other one will conduct for the opposite polarity of input signal. Since a relay coil is in the cathode circuit of each thyratron, one relay will be actuated for input signals of one phase while the other relay will be actuated for input signals of the opposite phase. Thus, if the motor is connected to the relays so that each relay

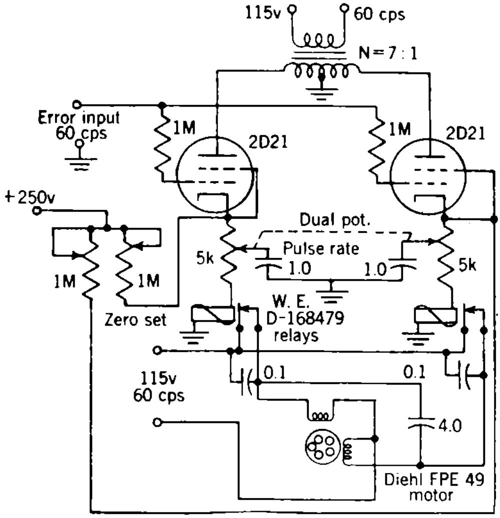


FIG. 12·65.—Simple thyratron and relay control for a-c motors.

controls one direction of rotation of the motor, the motor will be driven in a direction that is dependent upon the polarity of the input error signal. The pulse-rate control shown is an adjustment controlling the number of times per second the tubes conduct for a given amplitude of error signal.

The operation is as follows: The capacitor of the *RC* time constant that is placed in the cathode of either tube receives an increment of charge each time the tube conducts. This causes an increase in cathode voltage which prevents tube conduction for several cycles until the cathode voltage decays to a level that permits the grid again to fire the tube. Since the increment of charge per conduction cycle is nearly the same regardless of the size of the grid signal, for small grid signals the cathode voltage must decay for a longer period of time before conduction can again occur as compared with larger grid signals. Thus, the rate of pulses per second for small signals is less than that for large grid signals. The effect of this is to give a throttling action roughly proportional to amplitude of error, so that the rate at which the motor decreases the error is proportional to the error. Such a proportional response system is desirable, since it has greater inherent stability than a system whose rate of positioning is constant.

In addition, a control action approximating derivative control is present, for the lag of the cathode voltage with respect to an increasing or decreasing grid signal is proportional to the rate of change and hence introduces a bias term roughly proportional to the first derivative. For example, a fairly large, rapid increase in error signal will turn the system full on until the cathode condenser has charged up to the point where the normal throttling action operates. This is the behavior obtained with the more conventional proportional plus lagged-derivative types of controller characteristics. By adjusting the dual potentiometer, the time constant can be varied and the pulse rate adjusted for different applications.

In order to compensate for variations in tube characteristics, two zero set adjustments are provided. They are included so that a d-c bias may be applied to the cathodes to prevent tube conduction with zero input signal.

The wiring of the motor and relay contacts is self-explanatory if it is understood that only one of the relays is closed at any instant. The 0.1-μf capacitors across the relay contacts are included to protect the mercury contact.

Figure 12·66 illustrates a relay unit capable of higher positioning accuracy and faster response than that of Fig. 12·65. The error input device for which this was designed is shown for the sake of clarity, although other input devices may also be used. The servo motor is required to position the output shaft to correspond to the position of the shaft of the $5G$ input synchro. In the usual manner error voltage proportional to the sine of the angular difference between the shafts of the $5G$ and the $5CT$ synchros is developed at the rotor of the CT synchro. This voltage is series-added to a stabilizing voltage obtained from an a-c induction tachometer, and the sum is applied to the input stage of the servoamplifier.

Two stages of amplification are provided so that high sensitivity can be obtained for applications that demand high positioning accuracy. The output stage is similar to that shown in Fig. 12·65 with the exception that the relays are in the plate circuits of vacuum tubes. Again, the plates are fed with a-c signals 180° out of phase while the grid signals are in phase, so that each relay is actuated for only one phase of input signal. The connections between the motor and the relay contacts are identical with those of Fig. 12·65.

A feature of this unit that makes for good stability is the velocity voltage feedback. By employing an a-c tachometer generator as shown, a voltage whose amplitude is approximately proportional to the speed of the servo motor is injected into the input circuit. This is roughly equivalent to the addition of viscous damping to the output shaft. For further discussion of this subject, see Chap. 11. The ratio of velocity feedback voltage to error voltage can be adjusted for the optimum damping ratio by means of the potentiometer shown. A gain control adds flexibility to the unit.

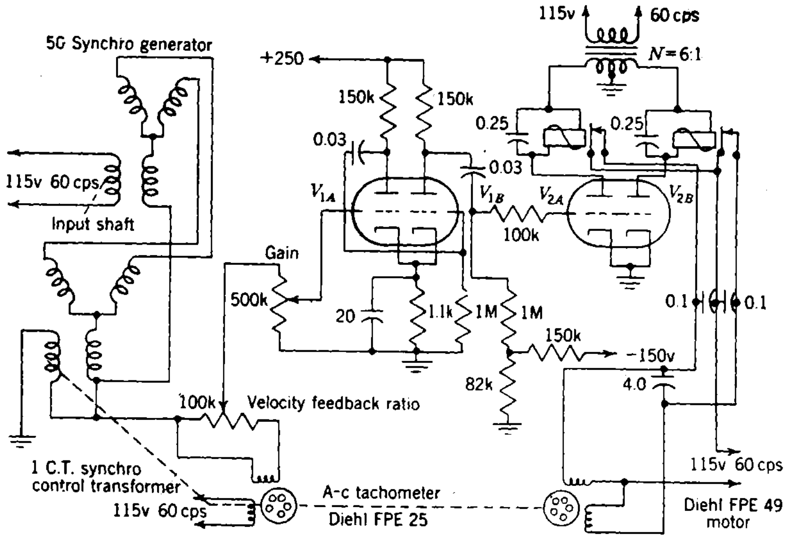


FIG. 12·66.—A-c tachometer-stabilized servo with relay control of a-c motor.

A third servo is shown in Fig. 12·67. This servo was designed to operate from 400 cps power. The motor is a Bendix CK-1, a two-phase induction motor. Usually, 26 volts, 400 cps are put across the low-impedance motor winding, and the output of a servoamplifier, phase-shifted roughly 90°, is applied to the high-impedance winding. The motor in this application turns a synchro-type resolver, described in Vol. 17, in such a way that the a-c voltage across one of the rotor windings is nulled. The error signal taken from this winding is amplified in two stages and then applied to the grid of the lower 6C4. A 400-cycle, 200-volt reference signal is applied to the grid of the upper 6C4. Consequently, the current through the two triodes and the coil of the relay is approximately proportional to the sum of the reference voltage and the amplified error voltage. This means that a signal of the same phase as the reference voltage will produce more than normal current, thus operating the D-168479 mercury contact relay, while a signal that is out of phase with the reference voltage will not operate the relay. This servo operates in a very unusual fashion. No provision is made for achieving proportional control. Instead, by using a very high gain, the hunt frequency is made high and the hunt amplitude low. In a linear system, this would be impossible, but nonlinearity here is such that the amplitude can be kept low. In addition, a viscous damper is used, geared down 30 to 12 from the motor shaft. The gear ratio between the motor and the resolver is 1000/1. With these features the servo hunts at approximately 15 cps

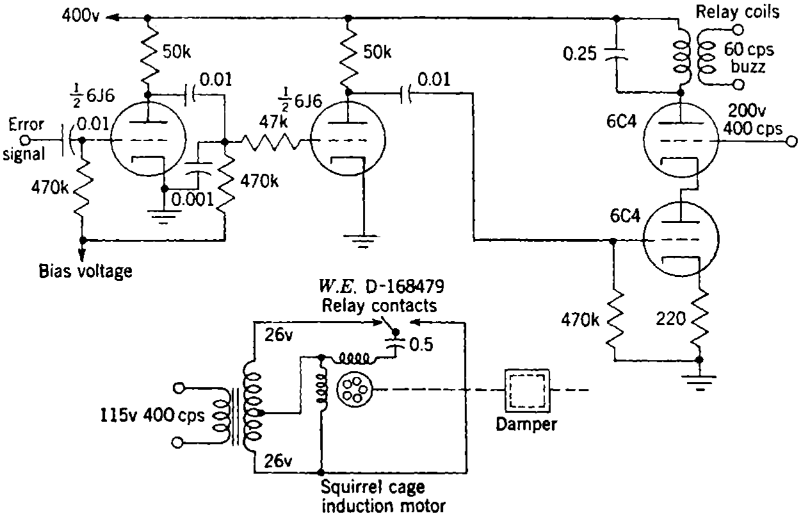


FIG. 12·67.—Relay servo for Bendix CK-1 motor. This somewhat unusual servo is operated at high gain, hunting steadily at about $\frac{1}{10}°$ amplitude.

with an amplitude of less than $\frac{1}{10}°$ measured at the resolver shaft. The success achieved in this instance with this simple but highly irregular method suggests that it could be used more frequently than it is.

12·25. Step, or Impulse, Motors.[1]—Step, or impulse, motors operate to give a fixed rotation of the output shaft for each input impulse. Some types may be controlled to give either direction of output rotation as desired; others must be installed in duplicate with the necessary means to energize the correct unit for the desired direction of rotation. Step and impulse motors are of particular interest for low cost, low performance, and servo drives or for those installations where inaccuracies will be averaged over a long period of operation and so cause no trouble. For exacting requirements, they are usually unsuitable.

Probably the simplest electrically operated step motor is the type with a solenoid-operated ratchet-type drive, so constructed as to advance the output shaft through a fixed angle each time the solenoid is energized. If drives in either direction are desired, two mechanisms are needed, and the operating pawls must be kept away from the drive wheel when not in use. Rapid wear of the mechanical parts may be expected under

[1] Section 12·25 is by J. R. Rogers and I. A. Greenwood, Jr.

continuous duty and is one of the major disadvantages of this type of design for use in servomechanisms.

The ratchet wheel and pawl construction is avoided in the "Dynatrol" step motor.[1] This inexpensive motor was used in motor-driven automatic radio receiver tuning. In this design, illustrated in Fig. 12·68, the ratchet wheels are replaced by a drum around which are mounted two solenoid-operated elastic belts to which cork friction surfaces have been cemented.

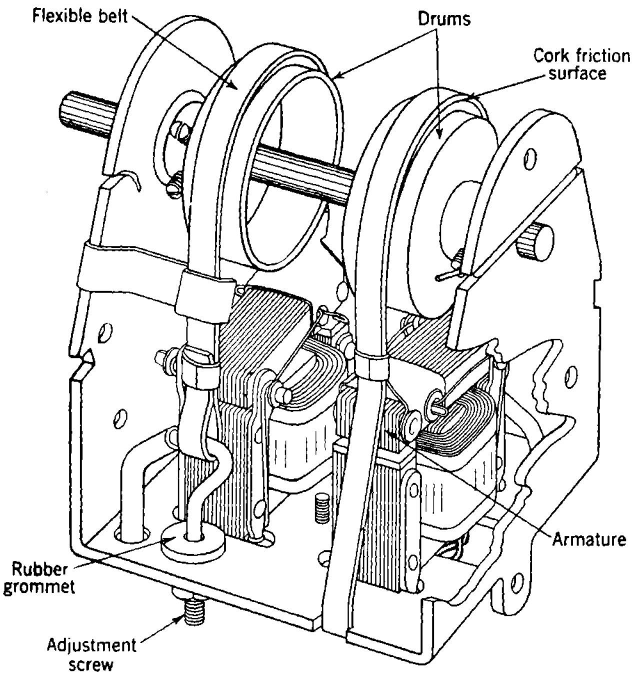


FIG. 12·68.—Crosley "Dynatrol" motor.

When either solenoid is energized, the band for the desired direction of rotation is tightened around the drum, rotating it by an amount determined by the solenoid plunger travel. When the solenoid is unenergized, the tension of the stretched belt returns it to its starting position. Inertia and friction of the drum on its shaft keep the drum from slipping back. As normally used, it is driven by applying 60 cps a-c power to the solenoid giving the desired direction of drive. The output is a series of small steps at 60 cycles equivalent for practical purposes to a continuous shaft rotation. The unit was incorporated in a contactor type follow-up with the power applied to the motor until it drives into a dead zone. This results in an extremely simple system but is adequate for the job required.

[1] Dynatrol is a trade name of the Crosley Radio Corp., Cincinnati, Ohio.

CHAPTER 13

EXPERIMENTAL TECHNIQUES

BY J. R. ROGERS AND I. A. GREENWOOD, JR.

13·1. Introduction.—The purposes of the experimental techniques to be discussed are (1) to obtain rough quantitative data on which to base a choice of components, (2) to determine accurately component characteristics for use in servo-system design, and (3) to provide experimental data on performance of over-all systems. The performance data will tell how closely the requirements have been satisfied and will also give a necessary check on the assumptions and values used in the analytical treatment, as well as checking the analytical computations. This chapter will discuss measurement first of components and then of system performance, in rough order of increasing complexity.

Available data on components adapted to servomechanism use are usually inadequate. Motor manufacturers, for example, publish data covering the normal operating conditions for their motors but they rarely give or have available data on performance under other operating conditions. As a result of the different source impedances, voltage and current waveforms, loading cycles, etc., encountered in servomechanism service, it is usually necessary to check motors with the proposed servomechanism drive circuit. The effective characteristics of motors may be drastically modified by feedback techniques; this fact also argues for checking with the proposed drive circuits.

Similarly, potentiometers and other input and output data devices are not completely described in available catalogues. Measurements of deviation from linearity (or from a required function), of resolution, of noise, of reproducibility, of temperature coefficients, etc., may often be necessary.

Amplifier design and the measurement of gain and phase characteristics are treated in Vol. 18 and in a number of textbooks.[1] For practical purposes it is usually sufficient to calculate an approximate gain for the

[1] F. E. Terman, *Radio Engineers' Handbook*, 1st ed., McGraw-Hill, New York, 1943, Sec. 5, pp. 947, 964–971; *Radio Engineering*, 2d ed., McGraw-Hill, New York, 1938, Chaps. V and VI; H. J. Reich, *Theory and Application of Electron Tubes*, 2d ed., McGraw-Hill, New York, 1944, pp. 156–164, 235–237, 665, 666; H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945, pp. 288–291, 479, 497.

amplifier, checking to determine whether and under what conditions this gain is actually obtained. Reduction of amplifier gain by such things as parasitic oscillations and electrical noise (e.g., tachometer ripple or motor brush noise) should be looked for experimentally.

Motor characteristics of interest may include armature inertia; static, coulomb, and viscous friction; speed-torque curves for various conditions of drive (supply voltage, error signal to the control circuit, and sometimes power-supply frequency); temperature rise of the motor windings and frame over ambient; and other characteristics for particular applications. As will be discussed, a simple set of parameters may be used to describe a motor for the purposes of engineering analysis of accuracy and stability. Not all of these need be measured to narrow the choice of a motor and control circuit to a few units, but complete data are essential for adequate design of the over-all servo system.

COMPONENT TESTS

13·2. Motors.—In analysis of a servomechanism for stability and accuracy, it is necessary to assign definite values to a small number of physical constants in order that a representation of the system be obtained which is suitable for the mathematical manipulations of ordinary engineering calculations. It has been found convenient to represent motors by the following three constants: *inertia, viscous friction, and a constant relating torque to applied current or voltage.* Although a very complicated series of nonlinear relationships would be necessary to specify exactly the characteristics of the usual motor, it is believed that these three simple constants represent a sufficiently good approximation to actual physical characteristics to satisfy the demand of most servomechanism engineering calculations. Reduction of the characteristics of a motor to three simple constants is possible by so choosing these constants that they include not only the magnitudes of the physical quantities to which their names refer but also modification as necessary in order to include the effects of other factors. While these three factors suffice for most of the purposes of stability and accuracy analysis, it is, of course, necessary to obtain data on a number of other characteristics of motors in order to engineer a servo system properly. It is usually found that other factors, such as static and coulomb friction and the effects of voltage fluctuations, frequency, temperature rise, etc., enter the calculations principally by way of modifications or limitations of range of the three constants listed above. Thus, for example, temperature data may be used to define the regions of the family of speed-torque curves that may be safely used in design. This section will first discuss the three constants listed above and then the experimental methods by which the data necessary for their calculations may be obtained. This will be followed by a discussion of other factors of importance in engineering a motor and control power circuit in a servomechanism.

The most important characteristic which may be modified in order to represent more than the physical quantity to which its name refers is the viscous damping constant,[1] usually designated as f or f_0. In the case of practically all motors used in servomechanisms, torque decreases as speed increases for a given applied voltage. The effect of viscous friction is to decrease the available torque by an amount that is directly proportional to speed. Because of this analogous behavior, the speed-torque characteristics of a motor may therefore be conveniently repre-

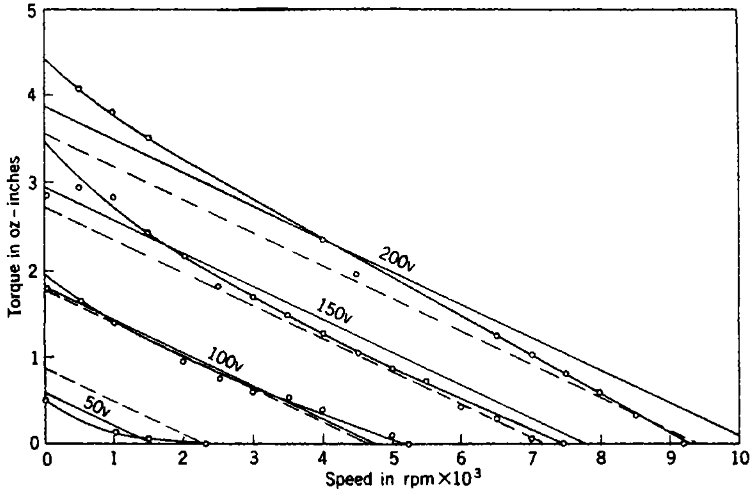


FIG. 13·1.—Speed-torque curves at various d-c control voltages for Elinco Midget type MS-1 motor. Broken straight lines show assumed linear characteristics.

sented mathematically by a suitable modification of the viscous friction constant of the motor. In the ordinary servo this decrease in torque with speed constitutes an important part of the servo damping. In some servos, this damping action alone may be sufficient for operation at low gain.

In choosing voltage-torque and viscous friction constants for a motor, it is convenient to plot torque as a function of speed for various values of applied voltage. A graph of this type is shown in Fig. 13·1 which represents the characteristics of an Elinco Midget MS-1 split-field motor. The average slope of the various curves is taken as the viscous damping coefficient. A constant relating stall torque and applied voltage must now be found such that the actual speed-torque curves will be approxi-

[1] The importance of this effect has been discussed by several authors; see, for example, H. Harris, "The Analysis and Design of Servomechanisms," OSRD-454, 1942.

mated by a series of straight lines of slope equal to the chosen viscous damping coefficient and with zero-speed (stall) torque intercepts given by the constant times the applied voltage corresponding to each of the respective speed-torque curves. One convenient method of finding the right voltage-torque relationship is to draw a best-fit set of parallel lines representing the speed-torque curves. The torque intercepts of these lines at zero speed may then be used, together with the control voltages of the curves that the lines approximate, to find the constant relating

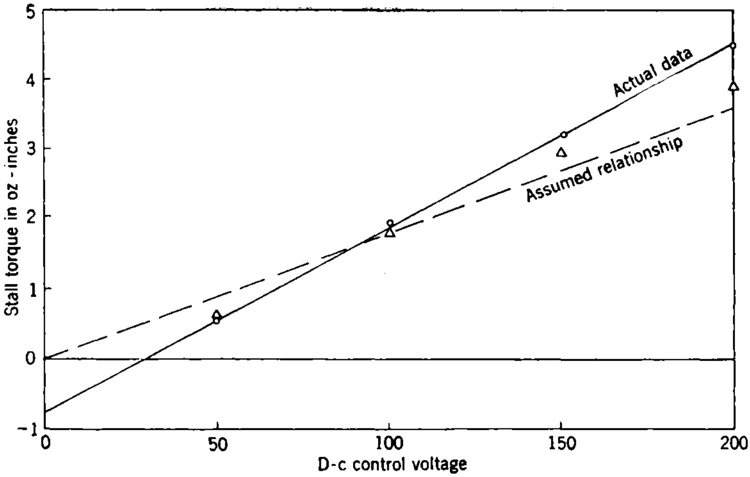


FIG. 13·2.—Stall torque vs. d-c control voltage. The broken line indicates the assumed linear characteristics.

stall torque and applied voltage. In Fig. 13·2, graphs of stall torque vs. control voltage are shown. In this figure, the solid line connecting the small circles represents the actual stall-torque vs. applied voltage characteristic, the small triangles represent the zero-speed intercepts of the best-fit parallel lines shown in Fig. 13·1, and the broken line represents the assumed torque-voltage relation. The broken lines of Fig. 13·1, whose zero speed intercepts are based on this relationship and whose slopes are the same as the best-fit set of parallel lines, represent a fair fit to the actual speed-torque curves over the working range of the motor.

In Fig. 13·2, the value of torque for the solid line corresponding to an applied voltage of zero may be taken as an indication of friction, either static or coulomb. The exact interpretation of this intercept, and in fact of any direct stall torque data, is complicated by the fact that friction may either add to or subtract from the torque produced by the motor, depending on the exact details of the experimental method used in taking this data. For this reason, it is preferable to measure coulomb and static frictional forces directly.

It will be recalled that static and coulomb friction are occasionally used in the theory presented in Chaps. 9 to 11. In view of this, it may be inquired why the assumed stall torque vs. voltage relationship of Fig. 13·2 is taken so as to pass through the origin when an obviously better fit to the triangular data points could be obtained by a straight line intercepting the torque axis below zero. The answer to this is that most of the equations of Chaps. 9 to 11 become unnecessarily complicated when account is taken of such a zero-voltage torque. For certain steady-state error calculations, however, static or coulomb friction, the equivalent of this zero voltage torque, may be taken into account.

From the curves it is thus found that this motor may be approximated reasonably well by the following assumed parameters: torque constant, 1.8×10^{-2} oz-in. per volt; and viscous friction, 3.8×10^{-4} oz-in. per rpm. The inertia of the motor need not be modified.

In many cases it will not be possible to obtain fits to actual data by linear approximation as good as those just discussed. In such cases it is best to fit the portion of the curves resulting in approximations that are conservative. Specifically, the torque constant should be chosen low, and the coulomb friction and viscous friction coefficient high. Choosing the viscous friction coefficient high leads to conservative estimates of velocity errors (*i.e.*, actual errors less than estimated errors) but optimistic estimates of the damping properties of the motor. In analyzing nonlinear devices such as motors on the basis of linear approximations, the extent of the approximations must always be kept in mind. The effects of nonlinearities are discussed in Sec. 11·11.

The method just presented may be compared with a method reported by Ferrell,[1] in which a constant relating stall torque and applied voltage and a constant relating zero-torque speed and applied voltage are measured and the ratio of these constants taken as the viscous term f (in his notation, R). Farrell's method has the disadvantage of not taking into account curvature of the speed-torque characteristics between their intercepts on the speed and torque axes and is also subject to possible confusion regarding friction when taking stall torque data. The latter disadvantage is not serious if careful technique is observed. The method, however, is somewhat faster and simpler than that given above.

Measurement of Inertia.—Armature inertia can be measured easily by the torsion pendulum method. A steel or other spring wire is fastened coaxially into one end of a cylinder of brass or other metal, of dimensions (and inertia) comparable to the armature to be tested. A hole in the other end of the cylinder should just fit the shaft of the armature in question. The torsional period T_1 of the brass cylinder alone is meas-

[1] E. B. Ferrell, "The Servo Problem as a Transmission Problem," *Proc. IRE*, **33**, No. 11, 763–767, November 1945.

ured; the armature is fastened into the cylinder; and the torsional period T_2 of the combined mass is measured. For a given suspension wire the periods and inertias are related[1] as shown by Eq. (1); J_1 is the moment of

$$\frac{T_1}{T_2} = \sqrt{\frac{J_1}{J_2}}, \tag{1}$$

inertia of the cylinder alone, and $J_2 = J_1 + J_a$, where J_a is the required armature moment of inertia. If the standard or comparison mass is a cylinder, its moment of inertia can be calculated from the formula $J = mr^2/2$. Usually the mass of material removed from the hole for chucking the armature can be neglected; if not, its inertia can be calculated and subtracted from that of the cylinder. It may be desirable also to calculate or to measure the inertia of the shafts in the gear train driven directly by the motor and to refer it to the armature shaft. To do this, divide the value obtained for each shaft by the square of the reduction from the motor to that shaft.[2] If the value obtained is one-fifth or less of the motor armature inertia, it may usually be neglected for practical purposes.

It is, of course, necessary to calculate or measure the load inertia as well as the motor and gear-reduction inertia. Often in instrument servos the load inertia reduced to the motor shaft is small compared with the motor inertia itself.

Torque.—A number of devices of all degrees of complexity are available for measuring torque. The field divides itself into those devices which measure torque transmitted and those which can produce a known or measurable torque loading. In the first class are calibrated spring couplings which may be viewed while rotating by means of stroboscopic illumination, piezoelectric devices, and resistance strain gauge devices which measure strain of some part under torsional stress. Indirect measurements of torque can be made by substitution methods, details of which will vary with the individual problem.

Typical torque loading devices are the prony brake, whose couple is measured by a spring deflection or weight lifted at a known distance, and the eddy-current brake or other electromagnetic dynamometer, which must be independently calibrated. One of the simplest and most practical methods, requiring no external measuring equipment, uses only a brass pulley with a flat groove, some cord, and a spring balance. Good quality linen fishline is very satisfactory for the cord, although almost any cord can be used if necessary. Unless a considerable length of the

[1] See almost any mechanical engineering handbook or college experimental physics manual for a discussion of this measurement.

[2] Section 11·12 further discusses the effects of gear reduction on inertia.

cord is in contact with the pulley surface, there is a tendency for it to grab at high tensions and low speeds, causing vibration of the spring balance and uncertain readings. For higher torques, a larger pulley should be used or an extra wrap of the cord around the pulley taken. Lubrication is unnecessary, but the pulley surface should be smooth. An arrangement of this type is suitable for torque measurements between a fraction of an ounce-inch and several foot-pounds. For the higher torque values other methods are generally to be preferred because the pulley-cord arrangement has limited heat-dissipating ability.

Friction brakes of the shoe type are illustrated in all the mechanical engineering handbooks. For instrument-type motors they offer few advantages other than larger power ratings over the spring-balance device just described, particularly as many of these brakes are designed for use with one direction of rotation only.

One useful design of torque loader combines a mechanical dynamometer torque-measuring device with a graphite on steel friction-plate torque loader. This device may be coupled directly to the unit to be loaded. It has the advantage of relative independence of speed and the ability to maintain its settings over a reasonably long run. Care must be taken to keep the friction surfaces clean.

An auxiliary motor may be coupled to the shaft to be loaded and so connected and energized as to oppose the rotation of the shaft. Calibration of the torque output against input power at the speeds of interest must be made, but once calibrated it is a convenient and useful test instrument.

Eddy-current brakes have certain advantages, particularly at low speeds. Like ordinary motors, they require calibration or the use of a torque-measuring device.

These torque-loading devices differ in their operating characteristics. The torque produced by the spring balance may exhibit "stick-slip" characteristics of serious magnitude at low speeds. A loader of the friction-plate type using graphite against steel can be made to have a torque load that is constant much closer to zero speed and does not have the same high starting friction. By use of a stiff spring and very sensitive displacement indicator such as a dial micrometer, this device can be made to have very little of the dead space or apparent backlash of the spring balance.

Because of the difference in characteristics of the various torque-loading and torque-measuring devices described, care should be taken in the choice of a torque-measuring means for attachment to an operating servo system. Fortunately, measurements of motor characteristics outside a servo system may be made much more conveniently.

Occasionally it is necessary to measure torque at low temperatures to determine lubricant and other loads. If an extension shaft can be run out of the cold chamber and mounted on ball bearings lubricated with very light oil, such as Univis 48, it may be coupled to the motor with a flexible coupling and the torque measurements made on the shaft extension. It is, of course, desirable to make a measurement on the shaft and its bearings with the motor disconnected but with all other conditions the same.

Speed.—Motor speed may be measured stroboscopically, by timing and counting the revolutions of a geared shaft, by measuring the voltage from a tachometer generator, by use of a spring drag device like the automobile speedometer, or by measuring the frequency of any waveform associated with the motor rotation. In practice, the first two methods are preferred for measuring motor characteristics. For low speeds (below 500 rpm) or for erratic speeds, stroboscopic methods are sometimes difficult, and it is easier to count (usually with a Veeder-Root type counter) the rpm of a shaft running at a known reduction from the motor. For speeds from 600 up to 14,000 rpm the General Radio Strobotac, model 631B, is satisfactory; it can be calibrated directly against 60-cycle line frequency and is easy to use.

A direct reading of the speed in rpm is given by the highest frequency setting at which a single distant image can be obtained of the shaft or gear being observed.

Special pattern disks can be attached to the rotating part which permit extension of the range of measurement to lower speeds. Twofold or fourfold symmetry will permit a proportionate reduction in measured speeds. The common stroboscopic disk for $33\frac{1}{3}$- and 78-rpm phonograph turntables gives an idea of the low speeds to which stroboscopic measurements may be extended.

If the stroboscopic light is run at 60-cycle line frequency, the motor shaft will appear to stand still at 3600 rpm, at 7200 rpm, and at higher multiples of the line frequency. At 1800 rpm the single pattern will appear double. Higher symmetry patterns will stand still at lower speeds. It is possible to design a stroboscope disk with several patterns so planned that one ring stands still every 100 rpm, another every 400, etc., making it possible to determine motor speed to a few rpm when the card is viewed under light from a stroboscopic source. Figure 13·3 shows such a disk.[1] When illuminated at 3600 flashes per minute, as by a General Radio Strobotac, successive rings appear to stand still at the following speeds, starting at the outside and going toward the center: every 100, 400, 600, 720, 900, 514, 1200, 720, 1800, and 3600 rpm. At 3600 rpm the entire card appears to stand still. If the illumination is at

[1] This is a copy of a tachometer test card described in *The General Radio Experimenter*, August 1943.

7200 flashes per minute as, for example, from a neon lamp flashed twice per cycle from a 60-cycle supply, these speeds will be doubled. Similar cards can be designed for special purposes and pasted on a pulley or gear. Where a Strobotac is not available, a 3-watt neon lamp with internal resistor, GE type Ne-40, for example, can be used from the 60-cycle supply, although care must be taken to prevent room illumination from masking the stroboscopic effect. It is not completely satisfactory, however, since the flashes are long compared with Strobotac flashes.

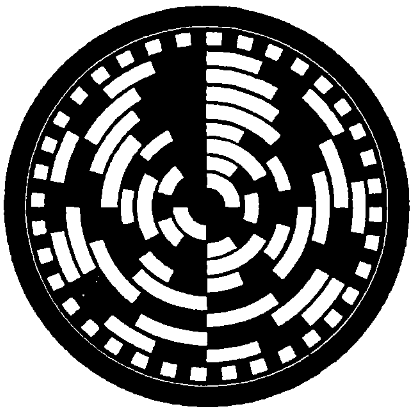


FIG. 13·3.—Stroboscope disk for speed measurements.

Another method of speed measurement makes use of a calibrated d-c or a-c tachometer generator whose output voltage (or frequency) is known as a function of speed. One must take into account the additional friction load imposed by the generator, which may be a considerable fraction of an ounce-inch for units commonly used. Once calibrated, a voltmeter can be adjusted to read speed directly and may result in an appreciable saving of time and effort when many measurements must be made. Where greater accuracy is desired, a precision potentiometer and null indicator (e.g., galvanometer) may be used to match the tachometer voltage, thus eliminating the meter errors. Measurements of speed may be made with tachometers to an accuracy of $\pm \frac{1}{2}$ to 5 per cent, depending upon the equipment used. It is convenient to measure power input to the motor at the same time speed-torque curves are taken. This gives enough data to determine power output and efficiency under all conditions. Power output may be accurately computed or estimated by the following approximate (6.5 per cent low) formula:

$$\textit{Horsepower} = \textit{speed in rpm} \times \textit{torque in ounce-inches} \times 10^{-6},$$

which is easy to remember and easy to use.

Measurement of Friction.—It is usually desirable to measure directly static and coulomb friction, since these measurements are extremely simple.

Friction may be caused by bearings, lubricant, brushes, or slip rings. The simplest way to measure running friction is to drive the motor through a calibrated spring coupling from another motor, observing the displacement of the coupling at the operating speed by stroboscopic illumination.

At low speeds friction can be measured by a spring balance or a pan of weights, a cord, and a pulley of known diameter mounted on the motor shaft. The cord is attached to the pulley rim, and weights put in the pan or tension applied through the spring balance to cause the motor to maintain a given initial speed without acceleration or deceleration. The product of the weight or tension required and the pulley radius gives the coulomb friction torque. Static friction (stiction) may be measured with the same equipment in an obvious way. Although providing a simple way to measure static friction, weights are difficult to use for measuring running friction, since it is not possible to average easily with the weight method, and because with this method a momentary increase in friction may stop the armature. With a spring balance, a slight added pull suffices to keep the armature turning slowly, and with a little practice the observer can learn to average the results or take the maximum as desired. Results should be easily reproduced to within 10 or 15 per cent with these methods of measurement.

Temperature.—Temperature rise can be determined in a number of ways. Surface temperatures can be measured approximately by a small mercury thermometer whose bulb is covered with putty to hold it in close thermal contact with the surface. Thermocouples can be installed in less accessible locations but require more equipment and are more difficult to use. If temperatures must be measured at several places on a motor, thermocouples are probably the best method. Surface thermometers are available from several manufacturers.[1] Such a device is placed in contact with the surface and gives a direct reading with only a few seconds delay.

Of greater interest than surface temperatures are the winding temperatures. The windings are not accessible, but fortunately an average value for the winding temperature can be calculated easily from the change in winding resistance between some known temperature (windings and frame at room temperature at start of test, for example) and at the temperature to be measured. The temperature coefficient of resistivity of copper wire[2] is approximately 0.004 part per degree centigrade.

[1] A representative instrument is made by Illinois Testing Laboratories, 420 N. LaSalle, Chicago, Ill., and is sold under the trade name Pyracon.

[2] Pender and McIlwain, *Electrical Engineers' Handbook*, 3d ed., Wiley, New York, 1936, pp. 2–04.

Motor Time Constant.—A simple and useful quantity specifying the dynamic response of a motor is known as the motor time constant T_m and is experimentally measured by observing the time required for a motor to reach 63 per cent of steady-state speed for any given applied voltage. The reciprocal of T_m is the angular frequency ω in radians per second at which the asymptote to the feedback transfer function on a decibel $-$ log ω plot breaks downward at 12 db/octave. For further details, see Chaps. 10 and 11.

13·3. Data Input and Output Devices.—One is interested first in the accuracy of input and output devices, such as synchros and potentiometers. Several other characteristics are also of importance in servomechanisms; this section will outline tests for these characteristics as well.

Synchros.—For a given pair of synchros for use in a servo, the departure of the control transformer position from that of the generator can be measured in degrees with relatively little difficulty. An accurate vernier-type dial is attached at each shaft. One unit is set to successive readings at desired intervals on the dial. With an oscilloscope across the output of the control transformer it is adjusted to the best possible null, and the angle is read. A plot of deviation against angle will show any periodic components as well as the limiting accuracy of positioning possible for a servo using the pair of elements tested, when connected as in the test. If one calibrated unit is used, the absolute accuracy of the other can be determined; otherwise the measured deviations apply only for the particular pair tested and only for the particular stator connection used. Alternatively, the two units may be coupled, adjusted to be at or near null, and the amplitude and phase of the CT output then measured for a number of angles. Care should be taken that any (flexible) coupling used does not introduce periodic errors in the relative position of the two rotors and that the shafts are not bent by a lateral stress due to misalignment of the rotors. For precise synchros where measurements must be made to a tenth of a degree or so, ordinary flexible couplings may introduce angular errors large enough to mask the actual instrument errors.[1]

Potentiometers.—Potentiometer linearity can be measured most conveniently by comparison with a linear potentiometer in a bridge circuit, using a microammeter, galvanometer, or oscilloscope as a null indicator. It is not difficult to arrange to display the curve of deviation on the oscilloscope screen as a function of rotation.[2] Special fixtures can also

[1] Satisfactory results can be obtained by using a coupling made of a $\frac{5}{8}$-in. diameter metal bellows with five or six convolutions. The end convolutions are soldered into collars bored to fit the shafts to be connected. This coupling has negligible periodic error; it can stand misalignments of the shafts as high as 5° and shaft offsets of up to $\frac{1}{32}$ in. If beryllium copper bellows are used, the units may be run at speeds up to 8000 rpm.

[2] See Chap. 18 and Vol 17 for detailed descriptions of such setups.

be devised to lock the standard potentiometer to the one being tested. For small quantities it is usually easier to use a dial on each unit and to compare dial readings at the null balance.

Most potentiometers, d-c generators, and other data devices are characterized by discontinuous output. For design purposes it is important to know the resolution, that is, the magnitude of the discontinuities. This subject is discussed in Secs. 12·1 to 12·8. Potentiometer resolution may be determined by counting the turns per unit of winding length and calculating the number for full scale. If a servo system has already been set up, as for measuring sealed units, it may be easier to use a sensitive voltmeter (perhaps coupled through a large condenser) to indicate the wire-to-wire increments of potentiometer voltage, counting the steps as the system is turned slowly through a measured angle. An oscilloscope with a fairly high gain amplifier may also be used for this observation.

Tachometer ripple may be conveniently observed with an ordinary oscilloscope or may be recorded on a recording oscilloscope for more careful studies of small irregularities and a permanent record.

Other Tests.—A characteristic of importance in many data input devices is the mechanical loading imposed by the device. Techniques applicable to the measurement of these loads may be essentially those discussed under torque measurements of motors in Sec. 13·3. Because of the very light loads involved, however, techniques must, in general, be more delicate. Pulleys, for example, should be very light. Thread should be used rather than linen cords.

The accuracy-measuring techniques discussed above in connection with synchros and potentiometers form the basic test for measuring the effects of electrical loading and for the interaction of parallel units. It is necessary only to repeat the simple tests described with the electrical loading or paralleled unit(s) present.

Other characteristics of interest affect amplifier performance and include waveform distortion, a-c pickup, potentiometer contact noise, thyratron transients, and motor brush noise. Observation on an oscilloscope of the error input to the amplifier or of the output of the data device will help to show trouble of this sort. In some cases it may be necessary to use a wave analyzer or sound analyzer to determine the frequencies present.

TESTS OF COMPLETE SERVO SYSTEMS

13·4. Inspection Tests.—The tests described above apply primarily to component parts of the servo system but do not cover performance as a system. The following sections outline quantitative tests of the over-all servo system, including methods of determining both static and dynamic characteristics. A number of simple inspection tests are available by which the servo designer may rapidly estimate servo performance and locate major trouble. This section will deal with such tests.

By turning the data input device through the angle necessary to produce motor rotation, first in one direction and then in the other, an estimate of the "dead zone" is rapidly obtained. A scale, preferably a vernier scale, on the data input device makes this an excellent quantitative test of servo performance under these conditions. An alternate method of measuring dead zone is turning the motor shaft by hand (keeping fingers away from gear meshes) and counting the number of revolutions between control action in one direction and control action in the other. This method is particularly convenient, since it utilizes the gear reduction between motor and data output device as a vernier. A fraction of the angle through which the motor turns must be attributed to backlash and must be determined and subtracted in order to determine the dead zone width as referred to the data output device. Counting the motor revolutions to reach maximum motor current gives a figure of use in velocity and/or torque lag calculations.

13·5. Quantitative Tests.—Positioning accuracy is determined mainly by the linearity and accuracy of the data input and output devices, accuracy of gearing, gearing backlash, width of control dead zone, component drifts, and sensitivity to temperature, line voltage, frequency, and other changes. Positioning accuracy is often referred to as a *static characteristic*. The measurements necessary to determine it are usually obvious; most of the special tricks have already been mentioned in Secs. 13·3 and 13·4. Accuracy and error are usually expressed in terms of per cent of full scale value; for some applications its absolute value is important. The application will determine whether maximum error, average error, or probable error should be specified. It is often helpful to give both maximum and average errors. Static characteristics are greatly affected by "buzz" or high-frequency superimposed oscillations, as discussed in Sec. 12·20.

A second set of characteristics, referred to as *dynamic characteristics*, describes the system response to changes in the input data. Included in this category are the various steady-state responses, the most important being velocity errors, acceleration errors, torque errors, and responses to sinusoidal steady-state excitations. Also included are the important transient characteristics.

Velocity error coefficient or *velocity lag* relates the error of a position servo in following a steady velocity input to the input velocity. It is expressed in degrees per degree-per-second or volts per volt-per-second, and its dimension is therefore just time. If specified in seconds no information is given as to the conditions of test; consequently it may often be preferable to specify the lag in degrees and the speed. Velocity error coefficient is measured by recording the speed as a function of input voltage with the loop open for inputs above and below the data output device setting, carrying the measurements up to the point of maximum motor speed. To open the loop, the data output device should be disconnected so that the motor can operate the load continuously without changing the error signal. The motor will run at a speed dependent upon the actual net error signal. This error is then the error necessary for the servo system to follow an input signal changing at the rate corresponding to the motor speed.

It is of some general interest to consider the increase of error caused by an increase of torque load and of great importance when the servo is to be subjected to large external torque loads. This increase may be directly calculated from the speed-torque curves of the motor and control circuits, together with a knowledge of the gain of the amplifier. It is measured in terms of error increase per unit of added torque load. It may be directly determined by measuring the error signal necessary to maintain a constant speed as torque load is varied. For servo systems whose motors can be approximated by the linear characteristics of the type shown in Fig. 13·1, the torque error constant is substantially independent of speed; however, increasing torque loads will mean lower top speeds and lower maximum torques available for acceleration.

Application of a step-function input to the operating system will give data from which damping characteristics and time constants can be determined. If there is available a curve tracer or recording oscillograph of adequate frequency response, these data can be most conveniently studied. The procedure consists of recording the servo output response and/or error following a sudden change in input data. If the system is underdamped, the amplitude of overshoot, the frequency of oscillations, and their decrement can be determined from the plot. Time constants of the system can be obtained, and magnitudes and characteristics of any positioning irregularities can be measured.

A curve tracer[1] will be found particularly useful for studying transient responses. It is capable of varying the position input data to the servo under test at constant rates and of switching rapidly from one value of input rate to another. A pencil or ink plot is made of servo error against input position. The unit can be driven out and back at constant speed; the difference between the error tracings for the two directions is then equal to twice the velocity lag for the speed of operation. Other data from this plot are dynamic fluctuations while following a smooth rate,

[1] Section 12·19 contains the circuit for the servo used with one such unit, the range-tracking calibrator.

roughness or noise or steps in the data output or reference element, and time for reversal at the speed of drive.

By switching the input rates, the transient response can be plotted to give damping, natural frequency of the system, and time constants.

For many servo systems data can be obtained with simple input switching and a slow-sweep low-frequency oscilloscope, particularly if the latter has a persistent-trace tube (P-7 phosphor such as 3BP7 or 5BP7).

Observation of the error amplifier output or of the data output device will give much of the information obtainable from a curve tracer. For d-c input servos, a slow-sweep generator capable of generating a d-c voltage varying linearly with time between the required limits will give helpful information on smoothness of following. Fixed voltages from a divider supplied from the same source as the data output device can be switched to provide convenient step inputs.

As discussed at length in Chaps. 10 and 11, introduction of sinusoidal waveforms into the feedback loop of a servomechanism will provide information on the resonant frequencies of the system, on damping, and on response to spurious signals and noise, as well as the expected input signals.

Valuable information is obtained from the response to sinusoidal excitation either with the loop closed or with the loop open. Response with the loop closed gives a direct measure of the frequency and amplitude of the resonant peak, thereby giving an indication of the "speed of response" and damping characteristics of the servo. Data taken with the loop open are important in the determination of modifications that can be made to improve stability and accuracy; in this case it is often desirable to measure phase response as well as amplitude response, although one is theoretically sufficient to determine the other for the minimum phase-shift networks used in servomechanisms. In the simplest technique for sinusoidal excitation one connects a small variable-frequency a-c voltage in series with either the input or output data device and observes servo response to this input. Oscillators suitable for this purpose are commercially available with frequency ranges from 2.0 cps upward.[1] Simple *RC*-oscillators can be made that will operate satisfactorily at frequencies down to a small fraction of a cycle per second. A beat-frequency oscillator using two high-stability *RC*-oscillators should be easy to construct for the frequency range up to 10 or 20 cps and would be equally useful.

Perhaps the simplest and most useful source of low-frequency excitation is a sine potentiometer, such as the RL 11-C,[2] turned by a small

[1] Hewlett-Packard model 200D.

[2] See Vol. 17.

geared-down motor. For very wide speed ranges this could be driven by a velocity servo; otherwise variation of motor voltage and/or change of gear ratios are usually sufficient to cover the range desired. This device has the additional advantage that it may be used to generate either an a-c or d-c synthetic error signal whose magnitude is varied sinusoidally at the low frequencies at which the system response is to be measured.

Motion of the servo output may be changed to a voltage suitable for observation and recording either by the data output device, if this happens to be satisfactory, or by an auxiliary potentiometer mounted on the output shaft or on a shaft intermediate between the motor and the output shaft. This potentiometer should be chosen for particularly high resolution, as it is usually of interest to take accurate data at frequencies where the output amplitude is extremely small. For those applications where the mechanical load of a potentiometer would be excessive, other data devices may, of course, be used.

In regard to sinusoidal steady-state data, a caution is in order. Sinusoidal steady-state analysis is based on the assumption of a linear system. If considerable nonlinearities are present and are entering the test data, such data are completely useless as far as this type of analysis is concerned. Practically, this means that amplitudes of sinusoidal excitation must be small enough so that all components are operated with nearly linear characteristics. A convenient test for this condition is to change the input excitation and observe whether or not the output changes proportionally. If there are system discontinuities such as potentiometer steps, their size will limit the minimum signal input for which a clean-cut, reliable set of data can be obtained. If it is difficult to observe amplitudes accurately for low signal inputs, this may indicate the need for better data devices or for cleanup of noise, microphonics, a-c pickup, or other disturbances. When taking open-loop data for transfer characteristics, it is essential that all elements work into the same impedances as with the loop closed.

There is much to be said for the practice of taking all sinusoidal response data with the servo loop closed, recording input, output, and error signal or only the last two. This has the advantages that conditions are not changed at all from normal operation and that the drifts and random effects which normally interfere somewhat with low-frequency open-loop measurements will cause less or no trouble. The feedback transfer function is the ratio of the output to the error, whether the loop is open or closed. If input, output, and error are recorded, both feedback transfer function and over-all response are available from the same test run.

There are several methods of measuring system phase shift under operating conditions. Phase shift, as a function of frequency, is most conveniently measured at the same time as amplitude response. Phase measurements are normally associated only with feedback transfer characteristics. If a d-c coupled amplifier oscilloscope is available, phase shift between the electrical synthetic error excitation and a voltage indicating the output shaft position can be easily determined. Phase shift can then be determined by the usual methods, for a number of frequencies in the range of interest.[1]

If a multichannel recording oscillograph is available, recordings of input and output waveforms on adjacent traces will give a permanent record from which phase shift can be measured directly.

Another method which may be useful is the use of an electronic switch and an oscilloscope capable of responding to the low frequencies of interest. Both input and output can be observed simultaneously, and phase shift observed directly. A motor- or vibrator-operated switch will suffice here. Mechanical comparison methods are possible, but the problem of correcting for varying amplitude of the output makes them relatively unsuited for phase-shift measurements.

A simple and useful method of obtaining data regarding the response-frequency characteristics of a servo system involves temporarily connecting an *RC* time constant of known value into the system and observing the change in the natural frequency. Since the effect of the time constant can be easily computed, such a measurement can yield information regarding the shape of the feedback transfer function.

[1] F. E. Terman, *Radio Engineers' Handbook*, 1st ed., McGraw-Hill, New York, 1943, p. 947.

CHAPTER 14

SPECIAL SERVO SYSTEMS

By J. R. Rogers, W. F. Goodell, Jr., I. A. Greenwood, Jr., D. MacRae, Jr., and J. W. Gray

14·1. AN/APG-5 Range Follow-up Servo.[1] *Description.*—This airborne servo system converts a d-c voltage representing target range, as measured by radar, into shaft rotation linearly proportional to the input voltage. This shaft rotation is used as the range input of a gunsight computer. A scale factor of 1.33 revolutions of the servo output shaft per volt change of the input is specified. The input voltage varies between +28 and +43 volts relative to ground. Direct-current power for the amplifier and for a motor-driven data output potentiometer are supplied by the radar system. The circuits require 10 to 12 ma of accurately regulated plus 250 volts and 2 ma from a VR-tube-stabilized minus 150-volt supply. In addition, the servo system requires approximately 65 watts of 115-volt, 400-cps power, regulation uncritical, and 10 watts of plus 26.5 volt d-c power for the motor field. For lower power consumption, a permanent-magnet field motor (Elinco type B-35 or PM-1-M) may be substituted for the Elinco B-64 wound-field motor, eliminating the need for the 26.5-volt supply.

The servo system consists of a two-stage d-c differential amplifier; a motor control circuit using type 2050 thyratrons; a gear box containing the wound-field motor, a potentiometer, and a power take-off; and the necessary intercabling. The amplifier weighs 5 lb, the gear box $3\frac{1}{2}$ lb, and the cabling junction box 3 lb. The servo system is integrated into the AN/APG-5 radar range-measuring system in such a way that the motor can be switched to manual control for emergency operation, a manual-automatic relay being located in the system junction box through which the various elements of the servo are cabled. Figure 14·1 is a photograph of the servo gear box, junction box, and amplifier. Figure 14·2 is a simplified schematic, indicating component values but neglecting the junction box and other cabling provisions. A number of points are of interest. The negative feedback network of Fig. 14·2 provides a convenient means of adjusting the servo gain as well as contributing to the servo stabilization. This network results in the negative feedback decreasing in magnitude with increasing frequency. The amplifier char-

[1] Section 14·1 is by J. R. Rogers.

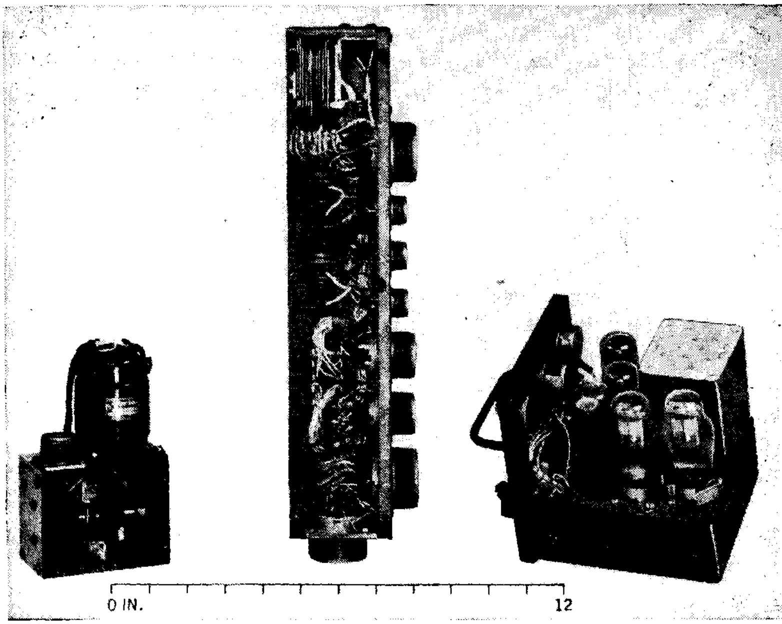


FIG. 14·1.—Gear box, junction box, and amplifier unit for AN/APG-5 range servo.

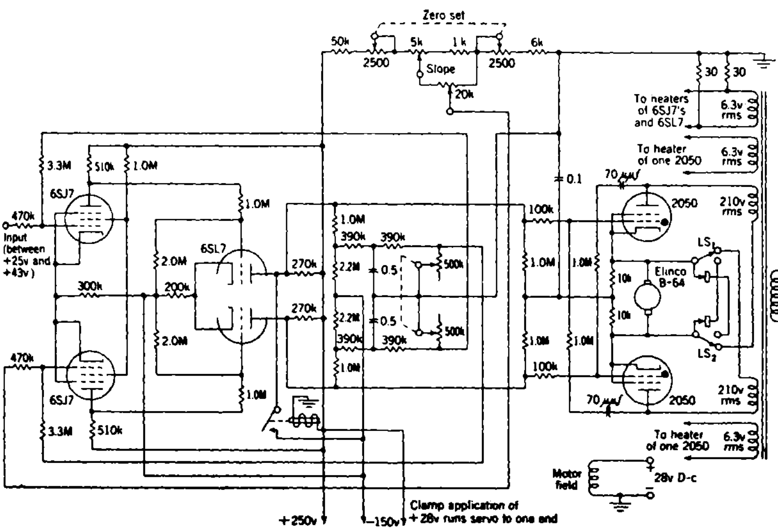


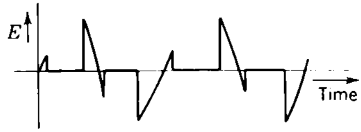
FIG. 14·2.—Schematic of AN/APG-5 servo.

acteristic thus obtained is equivalent to that of a phase-lead network and an amplifier with flat frequency response.

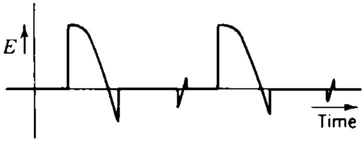
The characteristics of the thyratron motor-control circuit, described in Secs. 13·17 and 13·18, are such that smooth control of motor power is possible from zero to the maximum value and speed is roughly proportional to error signal magnitude up to the point of amplifier saturation. Figure 14·3 shows armature voltage waveforms with this circuit.

With armature voltage feedback as provided by this circuit, motor speed is roughly proportional to error amplifier output signal over a range of motor torque loads. A decrease of error voltage causes reduction of the firing angle for the next few positive half cycles or a complete reversal of power to the motor, even though the input error voltage may not have been reversed in polarity.

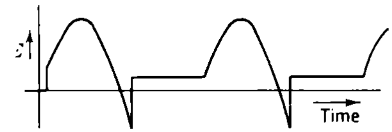
Provision is made for dynamic braking of the motor upon operation of the limit switches to reduce overtravel at either end of the allowable range. A selenium rectifier mounted in the system junction box is switched across the armature to provide what is essentially a short circuit for the generated armature voltage, thereby opposing continued rotation. Such a rectifier does not interfere with drive away from the limit switch, since in this case it acts as a high-resistance shunt.



(a) Motor at rest, positioning on stationary input, both thyrations firing for small angle.



(b) Same but one tube firing over larger and the other over smaller angle, corresponding to small error signal.



(c) Same but full error signal one tube cut off entirely, the other full on.

FIG. 14·3.—Waveforms of voltage across AN/APG-5 servo motor armature. (a) Motor at rest, thyratrons firing for small angles; (b) small error signal present; (c) full error signal, one tube off, other full on.

The data output device is a 20,000-ohm DeJur model 292 potentiometer chosen to have less than ± 0.3 per cent of full-scale deviation from linearity. A divider supplying direct current to this potentiometer is composed of all wire-wound resistors. The divider circuit permits of nearly independent adjustment of zero and scale factor. Voltage supply for this divider is obtained from the same regulated supply as the voltage to be measured, thus minimizing errors due to fluctuating line voltage.

Performance.—Static positioning error is 0.1 to 0.2 per cent of full scale. Nonlinearity of the reference element gives a maximum error of

± 0.3 per cent of full scale. With the normal gain setting to give slightly underdamped operation, positioning is smooth with average fluctuations of ± 0.05 per cent of full scale and maximum fluctuations of ± 0.15 per cent of full scale about the correct position, at all speeds up to 10 per cent of full scale per second. Maximum speed unloaded is 3.5 sec for full-scale travel, corresponding to approximately 350 rpm of the power take-off shaft. The recommended maximum connected torque load of 10 oz-in. on this shaft reduces the maximum speed to approximately 250 rpm. Stall torque is above 35 oz-in. at the output. Velocity error coefficient is approximately $\frac{1}{30}$ sec at speeds of 180 rpm.

The gear box is equipped with a thermostat and 28-volt heater which the thermostat turns on whenever the temperature of the gear box drops below $-16°C$. The gear box is lubricated with a low-temperature grease, and it represents a negligible load on the motor above this temperature.

The following criticism of the present system is intended to aid the designer in making use of this or similar circuits. The d-c amplifier drive to the control circuit requires that the motor and motor-control circuit operate at a potential of 125 volts above ground and requires that there be low leakage from the motor armature to ground. No trouble is encountered with the motors used, Elinco type B-64 or B-35, if they are fitted with straight carbon brushes and if the brush holders are so constructed (or modified) as to minimize the collection of carbon particles on the terminal boards. Carbon can collect across the terminal from the brush holders to ground or to the motor field terminal and cause leakage, making both thyratrons fire when the system is at balance. Current production motors have baffles installed around the brush holders, but some of the early units must be modified in order to give satisfactory operating life between brush-cleaning periods.

Any circuit in which the center tap across the motor is at ground potential will be less critical in this respect; for this reason a circuit with a-c amplifier and a-c error signal to the same control circuit is preferred. Such an a-c circuit has the further advantage that semipermanent readjustments of the tube elements under vibration or shock do not affect the calibration as they do with d-c amplifiers. This is of value for precise work where long-term stability of calibration is essential.

14·2. PPI Follow-up Servo.[1]—The plan position indicator (PPI) follow-up servo described in the present section mechanically rotates a sweep deflection coil around a cathode-ray tube in synchronism with radar antenna or other source of information. The major problem encountered in the design and construction of such a servo is the high degree of smoothness and accuracy required under conditions of large input velocities and accelerations and relatively high torque and inertial

[1] Section 14·2 is by W. F. Goodell, Jr., and I. A. Greenwood, Jr.

loads on the motor. The specific servo system described was designed for airborne use and is one of the best of its kind. With minor modifications in the phase-lead network and gear train, this servo may be used over a wide range of antenna speeds with a high degree of accuracy. Development of this circuit was a joint effort of the MIT Servomechanisms Laboratory, the General Electric Company, and the MIT Radiation Laboratory.

The operating specifications to be met are

1. Continuous antenna rotation at 6 rpm maximum.
2. Sector scanning[1] with maximum acceleration 90°/sec².
3. Torque load of approximately 20 oz-in. at a 10-speed drive shaft at 6 rpm antenna rotation speed.
4. Error of 0.5° maximum at constant velocity and 0.7° maximum when sector scanning.
5. Power frequency of 400 cps.
6. Compactness and lightness appropriate for airborne use.

The data input system consists of a Bendix AY-102-D synchro generator[2] on the flux-gate compass indicator (necessary for azimuth stabilization[3] of the PPI) driving an AY-131-D differential synchro on the antenna mount, which in turn drives an AY-101-D control transformer on the PPI coil assembly, this last synchro being the data output device. The over-all accuracy of the synchro system when used in this way is about $\pm 0.3°$. The entire system operates at one-speed, since this provides sufficient accuracy and eliminates extra phasing devices.

The electrical circuit for the amplifier and output stages is shown in Fig. 14·4. The rotor voltage of the AY-101-D synchro control transformer is fed to a phase inverter consisting of two halves of a 6SL7 triode and the balancing network R_4, R_6, and C_2. Resistor R_6 is made slightly larger than R_4 to compensate for the finite gain of V_{1B} and the loss of gain due to the divider action of C_2 and R_5. The plate and grid leads indicated are shielded to minimize pickup of unwanted power frequency voltages which might give rise to spurious signals. Capacitors C_3 and C_4 bypass to ground high-frequency pickup and high-frequency harmonic voltages which may be generated in the synchro system.

The a-c error signals are fed into a phase detector consisting of two halves of a 6SN7 triode V_{2A} and V_{2B}, two plate transformers T_2 and T_3, and the filter R_{12}, R_{13}, C_7, and C_8. The primaries of the plate transformers are connected to the 400-cps line which supplies the excitation for the synchro system. The operation of one side of the circuit will be

[1] See Vol. 6 for explanation of this radar term.

[2] *Cf.* Sec. 12·2.

[3] See Vol. 6 for explanation of this radar term.

described, operation of the other side being equivalent but of opposite phase. When the plate end of T_2 is positive, current flows through V_{2A} and R_{12}, charging C_7 negatively to provide a bias for V_3. The amount of current and hence the voltage to which C_7 is charged depends upon the potential on the grid of V_{2A} during this period. With no external error signal the common cathode resistor R_{11} provides a bias of -1.7 volts, which permits a current through the tube sufficient to maintain the bias on V_3 at -19.5 volts. If the error signal is in phase with

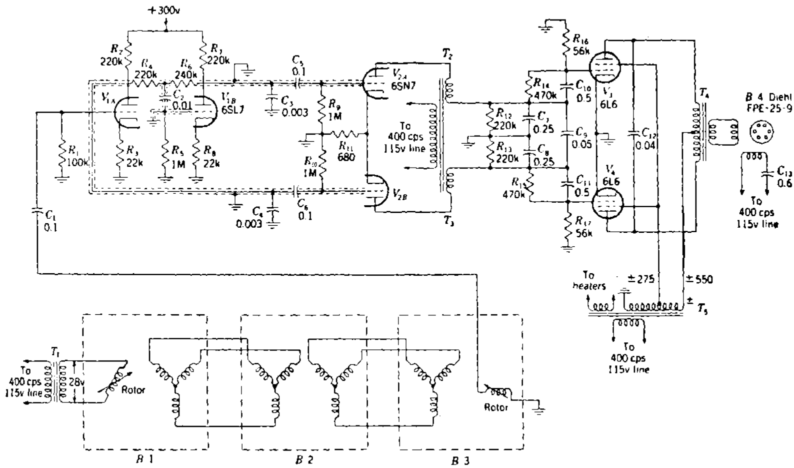


FIG. 14·4.—Schematic of PP1 follow-up servo.

the plate voltage, the average grid potential of V_{2A} during the period of conduction is more positive and the average current is therefore increased. This results in a more negative bias for V_3, and C_7 becomes charged to a higher negative value. If the error signal is 180° out of phase with the plate voltage, the average grid potential of V_{2A} during the period of conduction becomes more negative, the average current is decreased, C_7 charges to a smaller negative voltage, and the grid voltage of V_3 becomes more positive.

The above action is duplicated inversely by the other side of the circuit consisting of V_{2B}, T_3, R_{13}, C_8, and V_4 and so furnishes a push-pull input to the driver tubes V_3 and V_4. Since the operation of V_{2A} and V_{2B} is push-pull, no bypass condenser is needed for R_{11}. For an error signal corresponding to an angular error of 1°, the bias on V_3 rises to -13.5 volts, and the bias on V_4 falls to -25.5 volts.

Capacitor C_9 is placed across the output of the filter to balance the voltages applied to the driver grids. For filter output voltage components of frequency lower than 150 cps this condenser will charge and discharge, allowing an unbalance to occur on the driver grids. Very rapid changes of error signal such as might result from noise pickup will, on the other hand, not cause an unbalance on the driver grids. It is found that the error signal under normal operating conditions contains only negligible components of greater than 100-cps frequency.

The rectified and partially filtered error signals are then fed through a two-channel phase-lead network consisting of R_{14}, R_{16}, and C_9 and R_{15}, R_{17}, and C_{11}. This network stabilizes the servo in the manner discussed in Chap. 11. The two pairs of resistors also serve as dividers to scale down the voltages appearing across C_7 and C_9 to the correct values for the driver tubes.

The phase-lead circuit has an attenuation factor of approximately 20 db (10). This results in a maximum phase shift of approximately 55°, occurring at a frequency of approximately 3 cps. The magnitudes of condensers C_7 and C_8 are chosen sufficiently small so that with the plate impedances of tubes V_{2A} and V_{2B} in series with these condensers, the voltage across the condensers will be reduced by approximately 1 db at 10 cycles and approximately 6 db at 50 cycles, falling off at 6 db/octave at higher frequencies. This characteristic represents a compromise between the maximum filtering of the signal from tubes V_{2A} and V_{2B} and the minimum interference by way of introduction of lagging phase shifts in the servo response frequency range, zero to roughly 20 cps.

The driver stage consists of two 6L6 beam power tetrodes driving an output transformer push-pull. The secondary of this transformer is connected to one phase of a Diehl FPE-25-9 two-phase, four-pole, 400-cps, squirrel-cage induction motor, rated at 7 watts output. The other phase of this motor is connected through C_{13} to the common 400-cycle line. C_{13} is adjusted to give maximum output torque when an artificial error signal is introduced into the amplifier and thus compensates for any phase shifts that may occur in the synchro system or the amplifiers, as well as introducing the necessary phase angle[1] (nominally 90°) between the currents in the two motor windings. The plate and screen voltages of the 6L6 drivers are supplied from a transformer whose primary is connected to the 400-cps line. The phasing is such that the plates and screens of the drivers are positive and are drawing current at the time that V_{2A} and V_{2B} are nonconducting. This is done because during this time the bias voltages on the drivers are changing relatively slowly due to the smoothing action of the filter. When V_{2A} and V_{2B} are conducting, the filter is charging up rather rapidly through the low plate resistances of the tubes; but when they are nonconducting, the filter is

[1] *Cf.* Sec. 12·22.

discharging rather slowly through the much larger resistors R_{12} and R_{13}. This connection gives more stable operation.

If there is no error signal, the bias voltages on V_3 and V_4 are equal and the two tubes conduct equally. Since the plate currents oppose each other in the primary of the output transformer, there is no resultant voltage applied to the control phase of the motor. If an error voltage is introduced, the bias on one of the drivers will increase and that on the other will decrease, thus allowing more current to pass through one of the tubes than through the other. This unbalance causes a voltage to be applied to the control phase of the motor through the output transformer.

The motor used is a special Diehl motor designed for servo use, and its torque-speed and voltage-speed curves are the same for both directions of rotation. The motor has a rated speed of 11,000 rpm and a rated stall torque of 2.6 oz-in. The gearing is such that the motor runs at 1000 speed and therefore is normally operating at 6000 rpm. This allows sufficient excess speed and torque to follow the required accelerations.

The gearing in one of the models of the gear boxes is as follows: A small 18-tooth gear on the motor shaft drives a 72-tooth bakelite gear. This gear is made light to reduce the inertial load on the motor. The rest of the gears in the train are metal, as their effective inertia through the step-up of the train is not important. The shaft with the 72-tooth gear also holds a 16-tooth gear driving an 80-tooth gear. On the shaft of the latter gear there is another 16-tooth gear driving an 80-tooth gear. The shaft for this last gear is the power output shaft at 10 speed and is connected through precision worm gears to the deflection coil mounts. Also on this shaft is a 32-tooth gear driving a 64-tooth split spring mounted gear. This shaft again has a 16-tooth gear driving another 80-tooth split spring mounted gear. This final shaft is the one-speed output to which is coupled the AY-101-D synchro. The split gears are used to ensure as little backlash as possible between the one- and ten-speed shafts. All gears except the first 72-tooth gear are metal 32-pitch commercial gears.

The operation of the amplifier, motor, and gear train described is satisfactory for the prescribed use, that is, for following antenna rotation with a PPI sweep, the accuracy being somethat better than that specified. The servo will pull into position upon warm-up with very little hunting and will also follow the required velocity and acceleration smoothly.

14·3. Resolver Servo.[1]—The lightweight, low-power, position servo described in the present section was designed to position the rotor of an Arma resolver.[2] It is of special interest because of the variable loop

[1] Section 14·3 is by Duncan MacRae, Jr.

[2] Chap. 6 and Vol. 17, Chap. 10.

gain problem involved. This servo was designed largely by experimental methods. It uses a Bendix CK-5 induction motor and a saturable transformer[1] for power control and requires three miniature tubes. It will operate with errors not exceeding 0.3°, without the use of any controls; if a balance control is inserted, this accuracy is improved. Because of the way in which the resolver is used, accuracy criteria and performance specifications for the servo are, however, best expressed, not in terms of a fixed angular error, but rather in terms of a "position" error, as will be described in detail.

Problems of Resolver Servo Design.—A resolver having two stator windings that produce fields perpendicular to each other and two similar rotor windings may be used to transform rectangular coordinates to polar coordinates.[2] This is accomplished by impressing on the stator windings voltages that represent two rectangular coordinates x and y and by orienting a rotor winding with its axis perpendicular to the resultant magnetic field. The error voltage from the rotor winding is made the input to a servoamplifier, the output of which is connected to a motor that turns the rotor to a position of (nearly) zero error voltage, thus achieving the desired rotor orientation. The second rotor winding then has at its terminals a voltage r proportional to the magnitude of the vector sum of x and y.

In applications of this coordinate transformation to radar navigation and bombing computers, the error requirement usually specifies that the errors resulting from the circuit shall not exceed a fixed position error on the earth's surface, 100 yd being a typical value. A fixed angular error results in position errors that depend on the range of the point represented by the rectangular or polar coordinates; if the angular error is $\Delta\theta$, the distance error is $r\Delta\theta$. The quantity $r\Delta\theta$ is available as the servo error signal, since the error signal is proportional to both the angular displacement[3] from the position of zero error and the magnetic field produced by the vector addition of x and y, that is, to r. The result of this is that if in the circuit there is a source of error equivalent to the addition of a fixed voltage to the servo error signal, this corresponds to a fixed error measured as a distance on the earth's surface even though the angular error from this source will vary inversely with range. It is seen that this is a fortunate circumstance when the preferred method of error specification is in terms of a fixed distance error.

The fact that the error signal per unit angular error is proportional to r does, however, introduce a difficulty in the design, for this corre-

[1] Sec. 12·23.

[2] Chap. 6.

[3] More precisely, the sine of the angular displacement, but if $\Delta\theta$ is small, $\sin \Delta\theta \approx \Delta\theta$.

sponds to a variation of servo loop gain[1] with r. If the range varies by a factor of 20 (from 1 to 20 miles, for example), the loop gain will also vary by this factor. If at the lower limit of r the loop gain is made sufficiently high to overcome the effects of friction and of d-c level variations, there may then be enough gain at the upper limit of r to cause serious oscillations. This problem also arises in the case of other devices for converting rectangular to polar coordinates, such as sine potentiometers[2] used with servos.

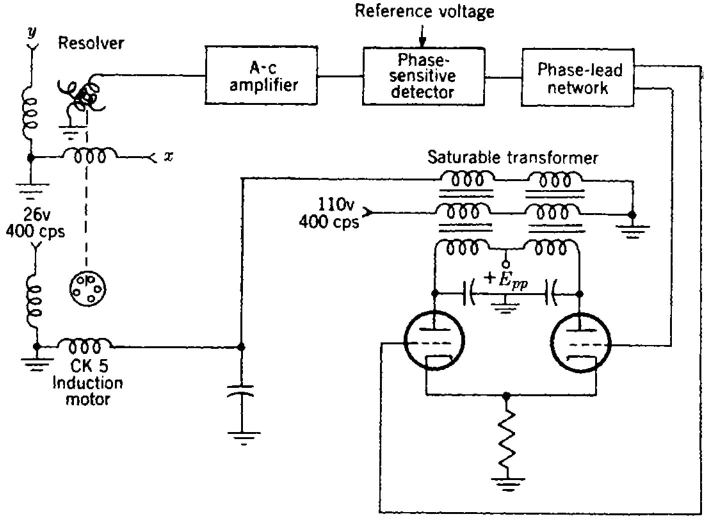


FIG. 14·5.—Detailed schematic of resolver servo.

It is possible to compensate for this variation in loop gain by the insertion of an automatic gain control if the variable r is available in the system. A straightforward but cumbersome method of doing this uses a servo to produce a shaft rotation proportional to $1/r$, with a potentiometer on this shaft used as a gain control for the amplifier. An accurate solution such as this is usually unnecessary. A variable-gain tube can be used in the amplifier with the r-voltage being rectified and used as a grid-bias gain control. The method used in the present design omits the automatic gain control altogether, in the interest of economy of parts. This results in somewhat inferior performance as compared with a constant-gain servo but nevertheless is satisfactory for the desired application. In the final design some oscillation is observed at large values of r.

[1] Variation of servo loop gain is discussed in Sec. 11·9.

[2] Chap. 6.

Design Procedure.—A schematic diagram of the circuit type chosen is shown in Fig. 14·5. An induction motor is controlled by a saturable transformer[1] which in turn is driven by the d-c plate current of two triodes in push-pull. This control circuit, as originally used by Bendix,[2] received a-c signals at the grids of the two triodes; an alternating reference voltage (the 400-cps line) constituted the plate supply for the triodes, which there-

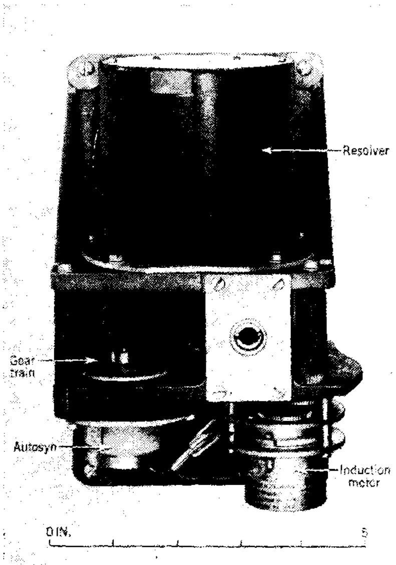


FIG. 14·6.—Arma resolver assembly.

fore acted as phase detectors. This method is modified in the present circuit in order that a phase-lead network may be inserted to reduce loop instability at large values of r. In general, phase-lead networks may be used with either a-c or d-c error signals; in the case of an a-c signal, however, the carrier frequency must be very nearly constant. The carrier frequency for which this circuit is designed, the output of a 400-cps aircraft power supply, is subject to variation of about ± 10 per cent, so that a-c phase-lead methods cannot be used. Thus a phase detector is used to make a d-c error signal available before the control stage. An a-c amplifier preceding the phase detector decreases the percentage effects of tube asymmetry and drifts in the phase detector.

The mechanical load of this servo is shown in Fig. 14·6. It consists of the resolver, a 500/1 gear train, and a synchro for transmitting the

[1] Cf. Sec. 12·23.

[2] Eclipse-Pioneer Division, Bendix Aviation Corp., Teterboro, N.J.

output angular information. A similar resolver servo unit using lightweight components is shown in Fig. 19·24.

The CK-5 induction motor is also shown in Fig. 14·6. When full power is supplied to the motor, it draws 0.2 amp in its 26-volt winding. In Fig. 14·5, a condenser is shown connected in parallel with the controlled motor winding; this has the functions of phase-shifting the fundamental component of the saturable transformer signal so that the induction motor will have a rotating field and of filtering the output waveform from the

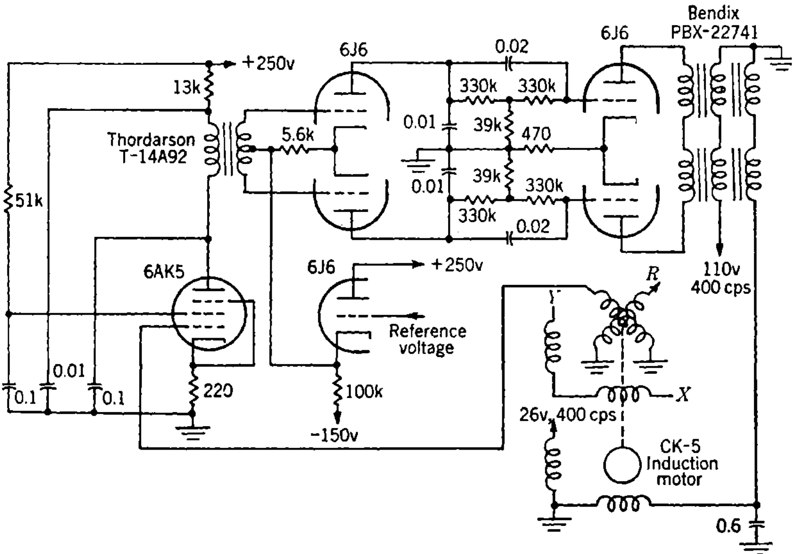


FIG. 14·7.—Detailed schematic of resolver servo.

saturable transformer. The phase shift produced by this condenser varies with frequency, however, so that the motor can be controlled satisfactorily only over a frequency range of approximately 350 to 450 cycles.

The saturable transformer used (Bendix Pioneer No. PBX-22741) requires approximately 2 ma of direct current to saturate one winding, in which case the filtered output from the secondary is about 60 volts rms. The d-c windings in the plate circuit of the triodes have many times the number of turns of the winding connected to the 400-cycle line. Therefore, even though the magnetic circuit is approximately balanced, 400-cycle voltages as great as 300 volts rms may be observed across the d-c windings when they are open-circuited. A serious limitation imposed on a circuit of this type by the use of a saturable transformer as a control element results from the time lag associated with the relatively large inductance of its d-c windings. This time lag, if large, tends to make the servo loop unstable. One method of reducing this time lag is to use pentodes as control tubes; the action is then more nearly that of current generators, and the reactor time constant L/R is accordingly reduced.

Miniature tubes, 6J6's and a 6AK5, are used in the design for compactness. The a-c amplifier uses the 6AK5 miniature pentode. A plate circuit transformer produces "floating" push-pull voltages for the phase detector, as shown in Fig. 14·7. The low primary impedance and high secondary impedance of this transformer make the gain from pentode grid to phase detector grids roughly 20.

The reference voltage for the phase detector is supplied from a cathode follower to avoid nonlinear loading of the source of reference voltage with consequent introduction of harmonics into the a-c computing voltages that supply x and y. The reference voltage allows current to flow on each half cycle when the cathodes of the phase detector 6J6 go below ground potential. Depending on the sense of the error signal, one or the other triode conducts more during this interval.

In each of the two output channels of the phase detector is a phase-lead network consisting of two 330-k resistors, a 39-k resistor, and a 0.02-μf condenser. Either the T-network of resistors shown or the equivalent π-network may be used. Resistance values of the T-network are lower, however, so that high impedances can be obtained without using high resistance values.

The performance of the system was checked by substituting a number of 6J6's in the circuit. The error $r\Delta\theta$ was observed on a calibrated oscilloscope screen with the aid of an amplifier. A voltage scale of 3 volts rms = 2000 yd was used. Systematic errors of $r\Delta\theta = 100$ yd were observed. At maximum range (40,000 yd) the servo oscillated with an amplitude of ± 270 yd, or $\pm 0.4°$. This could be observed easily on the oscilloscope. At a range of 20,000 yd the servo did not oscillate. The constant error of 100 yd then corresponded to an angular error of 0.3°.

14·4. Velocity Servos.[1] *Introduction.*—A velocity, or rate, servo is a servomechanism whose output speed, rather than position, is controlled by an input quantity, such as d-c or a-c voltage or current, the frequency of an electrical waveform, or a mechanical displacement. Many velocity servos may also be referred to as integrators, since the total output of a velocity servo is the time integral of the input quantity. This aspect of velocity servos is discussed in Chap. 4.

A velocity servo comprises a motor that drives the output, a speed-measuring device (tachometer) attached thereto, and an amplifier. The tachometer produces a quantity that is proportional to or otherwise a measure of the speed and is generally of a nature similar to the controlling

[1] Sections 14·4 and 14·5 are by J. W. Gray.

quantity. The amplifier receives the difference (error signal) between the controlling quantity and the tachometer output and controls the speed of the motor so as to minimize the error signal or some function of the error signal.

Most generally the control signal is variable, and the output speed is required to be an accurate measure of this over a fairly wide range, often including negative values as well as positive. In some cases, however, the velocity servo is simply a constant-speed device with a constant input signal. The control signal might not even exist as such, the error being developed by the action of the tachometer output on a special device such as a frequency-sensitive bridge.

Peculiarities.—The "error signal," unlike that of a positioning servo, is not necessarily a measure of the error of the servo. The error signal could be an appreciable part of the input signal; but if it were linear with respect to the latter and stable at any given speed, the device could be calibrated so that the signal would not matter.

If the speed is required to pass through zero, so that static friction reverses, or if the output torque is to vary at any given speed, the error signal will be subject to considerable deviation from linearity and stability. For high precision, therefore, the amplifier gain should be as high as is permitted by considerations of freedom from oscillation.

In the analysis and prevention of oscillation the same principles apply as in the case of positioning servos. Velocity of the motor takes the place of position, however, and acceleration replaces velocity. Thus, the rate of correction of an error is measured by the acceleration or deceleration rather than by the speed of the mechanical system. In this respect the prevention of rate servo oscillation is easier than in the case of a positioning servo, for while the mechanical inertia prevents any abrupt change of speed, it does not prevent abrupt changes of acceleration.

To clarify the significance of this difference, consider a simple positioning servo and a simple rate servo, each having amplifiers with no time lag and no corrective networks, which provide motor torque proportional to the error signal. Now, if a small abrupt change of input is made (not sufficient to drive the amplifier beyond its linear range), the feedback voltage of the rate servo, which is a measure of the velocity, begins immediately to approach the input. As it does so, the error and torque decrease, with a resulting exponential approach to equilibrium. The positioning servo feedback, on the other hand, being a measure of the position, which is the integral of velocity, will not immediately start toward the new input. Unless the viscous friction is high and the amplifier gain is small, at the time the error becomes zero the velocity is not zero, with resulting overshoot of position.

Another way of expressing the difference is in terms of the transfer locus from motor input to feedback voltage, as explained in Chap. 10. If there is no time lag between motor input and electrical torque, the transfer function is expressed as

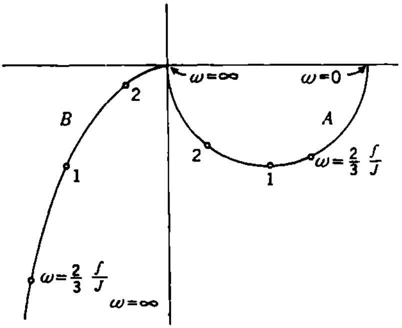


FIG. 14·8.—Servo motor transfer loci.

$$KG(s) = \frac{k_0}{Js + F},$$

rather than the expression of Eq. (10·5), since the feedback quantity is here proportional to $d\theta/dt$ instead of θ. This transfer function corresponds to a locus as shown by Curve A in Fig. 14·8 as compared with Curve B, which applies to the positional feedback. This difference allows more additional time lag to appear in a servo loop in the case of a velocity servo before oscillation will occur, as revealed by Nyquist's criterion. Just such an additional phase retardation may be imposed, however, by the necessity to filter the output of certain types of speed-measuring devices.

In many cases the performance with the amplifier saturated, i.e., full input power to the motor, is of critical interest. Some types of control operate normally in this condition, with the power alternating between full ahead and off or full reverse. In any case, saturation obtains if the input signal changes more rapidly than the available acceleration or deceleration. The quality that distinguishes the saturated behavior of velocity servos from the similar condition in most positioning servos is the lack of symmetry in the former—deceleration is there greater than acceleration. The two curves of Fig. 14·9 show how the speed may change after a complete reversal of power one way or the other. The abrupt change in rate at zero speed is due to the reversal of static friction. If the servo is operating in the neighborhood of a given speed, as that of the dashed line, the rate of error correction with the amplifier saturated in one sense will be measured by the slope of Curve A where it crosses the dashed line, while with saturation in the opposite sense the slope of Curve B applies.

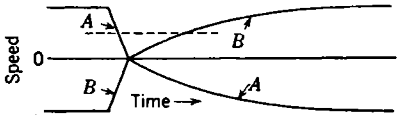


FIG. 14·9.—Typical acceleration curves at rate servo motor with load.

This peculiarity of rate servos may cause considerable error in a simple power-reversal type of servo even though the amplifier sensitivity is very high. In Fig. 14·10 the distance from the input level to either dashed line represents the error required to cause a power reversal. But after the speed, as measured by the tachometer output, has crossed either of these limits, a certain time lag ensues (e.g., relay closing time or motor field inductance effect) before actual reversal occurs. The difference between acceleration and deceleration results in the average speed being below the input signal, as shown, and this difference will be a function of mechanical load.

Types of Velocity Servos.—The choice of motor and control circuit for a velocity servo is largely the same as for positioning servos and depends on various factors, such as available power sources, requirements as to mechanical load (including tachometer), acceleration, speed range, precision, smoothness of operation, etc. If the range of speeds embraces zero and if precision and smoothness are desired, the motor should be as homogeneous as possible with respect to rotation; i.e., it should have the minimum of "slot lock" due to magnetic saliency.

The form of the input signal is usually predetermined by the kind of service to be performed. The input narrows the choice of tachometer, method of comparison, and amplification of the error signal. Types of tachometers suitable for use in integrating rate servos are discussed in Chap. 4 and in Sec. 12·6 and include those developing d-c voltage, d-c current (condenser tachometers), a-c voltage, and mechanical displacement. Other types of tachometers, which are applicable in rate servos for purposes other than integration, include those whose frequency output measures the speed and nonlinear devices like centrifugal governors, which operate within a small speed range.

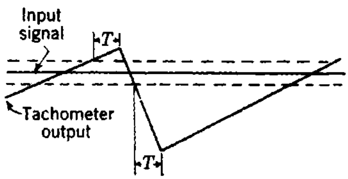


FIG. 14·10.—Action of power reversal in rate servo with time lag.

Production of Error Signal.—In the case of a voltage control and a generator tachometer the error signal may be developed by means of a simple series addition, as both generator terminals are generally available. In some circuits parallel resistance adding may be more convenient, although this reduces the error scale factor. If a commutated condenser tachometer is used in conjunction with a d-c voltage input, the speed is made to be such that the current passed by the device is sufficient to produce across a resistor a voltage equal to the input.

There are a large number of possible ways to develop electrical error signals from mechanical-output tachometers. The simplest is a contactor between the controlling position (which may or may not be variable) and the tachometer output, which switches the motor power on or off (or reverse) either directly or via a relay. Other possibilities include photoelectric pickoffs, capacity-change devices, flux-linkage devices, and electric strain gauges. Any method that gives "proportional" error signal, i.e., a signal that changes continuously and roughly proportionally with the error, is capable of greater precision than one that does not, for the reason illustrated in Fig. 14·10. This may be achieved even in the case of the contactor type, by superimposing on the normal contactor motion an excitation that causes rapid opening and closing in the vicinity of zero error, in such a way that the ratio of time closed to time open varies continuously with error.

A frequency type of tachometer may be an actual synchronous generator, a simple contactor on the motor shaft, or simply the use of some electrical ripple characteristic of the motor. The input control frequency may not exist as a voltage or current, being instead a characteristic of some network, such as an *RC*-bridge, upon which the measured frequency is impressed and which develops the error as an output voltage. If there is an input wave whose frequency is to be followed, the error might be derived either from the frequency difference or from the phase difference. The latter is, of course, much more precise but has the difficulty that if an appreciable frequency error ever exists (e.g., following a transient of some kind), the algebraic average error signal during each "slip" cycle is zero, so the operation is incorrect. It is possible to add the phase error and frequency error together so as to obviate this difficulty. A frequency error may be developed as the difference between the rectified output of two equal frequency-sensitive networks which receive respectively the control and the measured waves, these having been equalized in amplitude by the use of AGC. Another possibility is to run a synchronous motor from the input wave and compare its position with the output rotation.

Amplifiers.—The output or power stage of the amplifier is dictated by the type of motor and the available power. Various types are discussed in Chap. 12. In general, smooth control of the motor is preferable, and the rate of response, in terms of change of accelerating torque, should be as great as possible.

The voltage amplification is largely affected by the same considerations as in the case of positioning servos (*cf.* Chap. 12). Where the error is a d-c voltage, the simplest amplifier is direct coupled, but this is subject to considerable drift, and in certain applications requiring high stability a modulator or "chopper" is used, with a-c amplification.

One very useful device, which is also occasionally applicable to positioning servos, is the division of the amplifier into two parts, with *RC* feedback over the first part, as described in Chap. 4. This serves a double purpose: It stores, or "remembers," the integral of any large error that may accrue during rapid acceleration, causing this to be canceled by a measured overshooting of the speed, and it affords an excellent means of damping and inhibiting over-all oscillation. The memory feature operates by Miller feedback integration and is explained in

Chap. 4. This feedback also prevents any appreciable amount of the tachometer commutator ripple from appearing at the grid; such a ripple could decrease the effective amplifier gain by saturation.

The damping action is that of integral control, as explained in Chaps. 10 and 11. If the part of the amplifier over which the *RC* feedback exists has infinite gain by itself (as, indeed, it may have if positive feedback is employed), its transfer locus with the feedback is as shown in Fig. 10·20 for the "ideal integral controller." This locus is multiplied by that of the rest of the servo loop, giving a very high loop gain at low frequencies and still avoiding the point of oscillation as the frequency is increased.

14·5. Examples of Velocity Servos. *Thyratron Rate Servo.*—A medium-precision motor-control system of extreme simplicity, yet having all of the aforementioned elements of a rate servo, is shown in Fig. 14·11. The motor has constant field excitation or a permanent-magnet field, so that when the thyratron is not conducting, the armature voltage is proportional to its speed. The error signal or difference between this and the control voltage appears between cathode and grid. Whenever the armature voltage drops to within a volt or two of the grid potential, the thyratron will conduct on the next positive swing of the plate slightly accelerating the motor. Conduction ceases when the plate drops below the cathode, and the armature voltage again is a measure of speed. Thus, at equilibrium, voltage produced by the motor while it acts as a tachometer is slightly above the input control voltage by an amount depending on the thyratron characteristics.

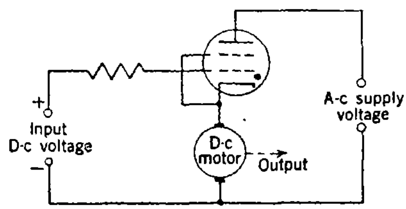


FIG. 14·11.—Thyratron rate servo employing back emf as tachometer.

The erratic effect of commutator ripple upon the firing of the thyratron may be objectionable in some applications; it may be eliminated or reduced by an appropriate condenser across the motor.

If it is desirable that zero speed correspond with zero input, a small amount of a-c voltage of the proper phase may be added at the grid. This may also be used to obtain a slower minimum speed, by permitting firing late in the positive half cycle and thus reducing the energy delivered to the motor. Ratios of minimum to maximum speed of the order of 1 to 500 are obtainable, and the linearity may be as good as 1 per cent between control potential and speed.

This is a unidirectional rate servo, but if the control potential is from a rectified a-c voltage, a reversing relay may be employed similar to that described below. On the other hand, if the input is alternating voltage

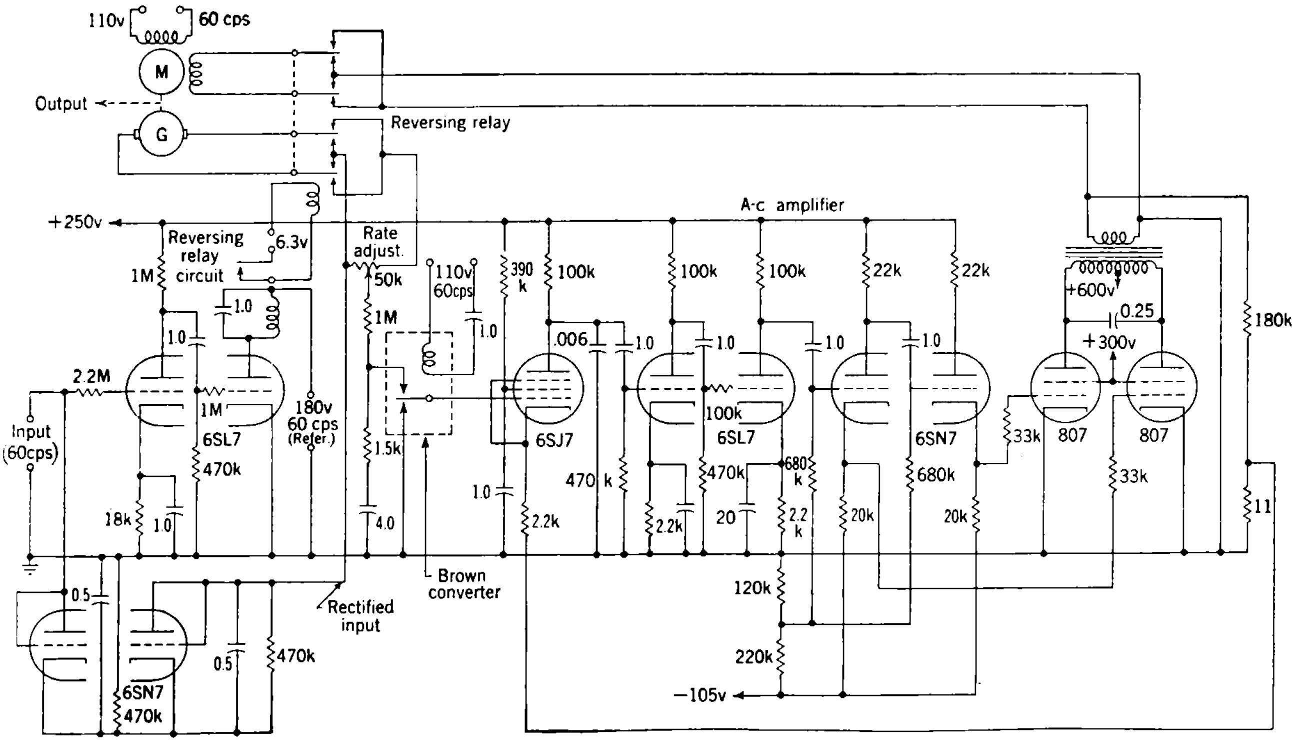


FIG. 14·12.—Rate servo employing converter and a-c amplifier to drive an induction motor.

of the same frequency as the power source, and if unidirectional operation is satisfactory, the input need not be rectified. With the input voltage roughly in phase with or somewhat leading the power voltage, the speed will be such that the back emf is about equal to the peak of the input.

If the motor has a wound field, the field current should be made independent of temperature by the use of a constant-current circuit or by compensation. If a fixed d-c voltage is employed across the winding, the field current will decrease with temperature rise, at about 0.4 per cent per degree centigrade, with an equivalent speed increase resulting. Permanent-magnet fields also have temperature coefficients; for example, 0.03 per cent per degree centigrade for Alnico V.

Rate Servo with A-c Amplifier.—Figure 14·12 shows a much more elaborate rate servo, used in radar trainer equipment, in which the error voltage is converted to an a-c error voltage by a vibrator (such as a Brown Converter), amplified as such and used to control a two-phase induction motor, whose speed is measured by a tachometer generator. A rectifier and a switching circuit are provided, permitting the use of an a-c control potential if desired.

The relay circuit is phase sensitive, causing reversal of the motor when the a-c input becomes negative with respect to the reference. Tachometer output must be reversed as well.

The rectifier is of the balanced type, having an auxiliary diode in series with the detector output, to cancel the effect of heater voltage variation and to produce zero d-c output at zero a-c input (assuming equal diode characteristics). The rectifier output is the rate servo d-c input, from which the tachometer voltage (or a certain fraction of it) is subtracted to produce the error voltage.

The error voltage is converted to a 60-cycle square wave by the Brown Converter. This is amplified by a multistage Class A amplifier having over-all negative feedback to reduce its output impedance. The output drives one phase of an induction motor, whose other phase is excited from the 60-cycle source. These voltages are in quadrature by virtue of a phase shift at the converter. The amplifier gain is about 15,000 (84 db) with feedback, so the error voltage need be only a few millivolts for full power to the motor.

The resistances and condenser at the converter input attenuate the high frequency ripple of the generator by a factor of 1000/1.5, without attenuation of the d-c error voltage. The 1.5-k resistance preserves a small amount of a-c gain without phase retardation—enough to keep the time lag of this filter from causing oscillation. The stabilizing action is that of integral control, as explained in Chaps. 10 and 11.

Rate Servo with RC Feedback.—In the rate servo of Fig. 14·13, a direct-coupled amplifier is employed to drive a pair of relays that control the power delivered to a small 28-volt motor. Smooth or "proportional" control of the motor is obtained by vibrating the relays at 60 cps by use of auxiliary coils (Sec. 12·20). A continuous change of power to the motor is obtained, from full power in one direction to full power in the reverse direction, as the input E_p to the relay control tube is moved through a range of about 7 volts.

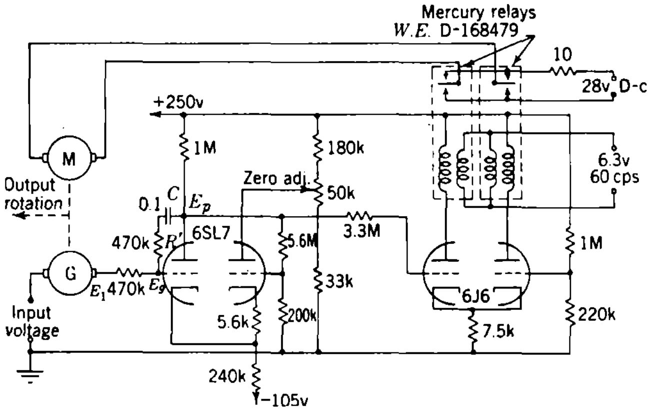


FIG. 14·13.—Rate servo employing RC feedback for damping and error integration.

The first stage of the amplifier has two triodes arranged so as to cancel the drift caused by heater voltage variation (*cf.* Vol. 18). Positive feedback is provided to the second grid, such that the first grid voltage E_g will remain within a few millivolts of zero as E_p moves through a considerable range.

The network between E_p, E_g, and the error input E_1 prevents the first stage from being saturated by commutator ripple or sudden changes of input, stores any error at E_1 which might result from inability of the mechanical system to follow rapid input changes, and prevents hunting. These functions are explained in Chap. 4. This type of network was apparently first reported for use in velocity servos by F. C. Williams.[1]

Pitot Marine Log System.—The pressure produced by a centrifugal pump when the water is not permitted to flow bears the same relationship to the speed of the pump as the pressure from a pitot tube bears to its speed through the water. Thus if the pump pressure and pitot pressure are made to be equal by adjusting the pump speed, the latter will be proportional to the pitot speed, and total rotation will be a measure of distance traveled. A rate servo may be used to correct the pump speed on the basis of difference-of-pressure data.

[1] F. C. Williams, "The Velodyne," TRE Report No. 4035.

Such a system[1] is shown schematically in Fig. 14·14. The comparison of pressures is done by a bellows, exposed internally to the pitot pressure and externally to the pump pressure. If the former exceeds the latter, the bellows expands and pushes a contactor which closes the circuit of one of the shaded poles of a single-phase induction motor, causing this to turn a variable auotransformer so as to increase the power delivered to the pump motor, thus increasing its speed. If the pump pressure is too great, the opposite action takes place. In the vicinity of equilibrium between pressures, the contactor is caused to vibrate between the two

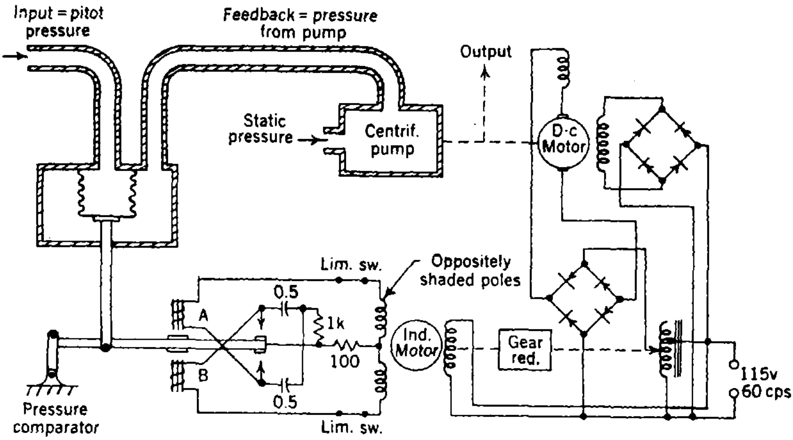


FIG. 14·14.—Ship's log using rate servo.

positions, by the magnets A and B, each of which tends to pull the contactor arm away from the contact that is made. This rapid vibration gives a "proportional" control characteristic to the error input and counteracts an over-all tendency to hunt. Limit switches in the shaded pole circuits are mounted on the variable autotransformer to prevent the induction motor from overdriving.

The cascading of time lags as a consequence of the means used to achieve a high power gain is stable in this velocity servo, but a similar cascading of time lags in a position servo would necessitate considerably more in the way of stabilizing circuits or devices. The explanation has been given in Sec. 14·4.

[1] This type of log is made by the Pitometer Log Corporation. A somewhat similar device for aircraft, using air instead of water, is called an "Air Mileage Unit" and is made by the Eclipse-Pioneer Division, Bendix Aviation Corporation.

PART III

VOLTAGE AND CURRENT REGULATORS

CHAPTER 15

REGULATOR ELEMENTS

By A. Jacobsen and J. V. Holdam, Jr.

15·1. Introduction.—In the design of electronic instruments it is often necessary to maintain voltage or current constant within specified limits or to have voltage or current determined within specified limits by some controlling voltage, current, or mechanical displacement. To achieve this result devices known as voltage or current regulators are commonly used.

Regulators may be classified as "simple" or "degenerative." Degenerative regulators depend on a negative feedback loop to provide power at low impedance in voltage regulators and at very high impedance in current regulators. Simple regulators do not use feedback but simply combine nonlinear elements with linear elements to effect a low output impedance for voltage regulators and a high output impedance for current regulators.

Most degenerative regulators contain these basic elements: a reference element which may be a voltage source, a current source, an electromechanical device in which a spring or mass is used for reference, or a nonlinear element in which a small change in voltage or current is realized for a large change in current or voltage; a sampling circuit yielding some desired function of the regulator output; a comparison circuit to provide an error signal representing the deviation of the output from the output desired; and a control element. These divisions are convenient because they are clear-cut for a number of the most important regulators, but in others two or more of these functions may be performed by the same element.

Factors of importance in specifying the performance of regulators include voltage, current, and power-handling capacity; impedance of output; output ripple; range of inputs and outputs; long-time stability; short-time or dynamic stability; absolute accuracy; and frequency characteristics.

Long-time Stability.—Long-time (static) stability may be defined as the constancy over a period of hours of the output averaged over periods of seconds. Long-time stability is impaired principally by slow changes in the characteristics of the sampling circuit, the comparison circuit, or the reference element. The extent to which these effects must be eliminated depends upon the performance required of the regulator.

Dynamic Stability.—Short-time, or dynamic, stability is defined as the constancy of the instantaneous output. Dynamic stability may be impaired by noise or microphonic effects in one of the elements or by inadequate stabilization of the degenerative loop.

A degenerative regulator may be considered as a negative feedback amplifier for which the input function is a constant. The theoretical principles for the design of stable high loop gain degenerative loops are presented in Vol. 18 and are similar to the principles of servo stabilization treated in Chaps. 9 to 11 of the present volume. See, in particular, Sec. 10·4.

15·2. Reference Elements.—Reference elements can be divided into three broad classes: constant-voltage elements, constant-current elements, and networks containing nonlinear impedances.

Constant-voltage Elements.—Sources of emf that are nearly constant with time are used as reference elements for regulators. The most important characteristics are constancy of voltage with aging, change of temperature, current drain, vibration, and change of position. Other characteristics important in the practical use of a reference element are voltage magnitude and effects of momentary short circuits.

Batteries.—Dry batteries can, for laboratory purposes, provide a voltage reference that is accurate to about 0.05 per cent for a period of several months when operated under specified conditions. The important requirements are fresh batteries, operation at all times with very low current drains (less than 1 μa), maintenance of constant temperature (within a few degrees), and maximum temperature not much above 25°C. Low temperatures do not damage dry batteries; in fact, their "shelf life" materially improves at 0°C. High temperatures (50° to 70°C) materially reduce the life of batteries and hence should be avoided. The temperature coefficient of 45-volt Z30N small Burgess batteries is about 0.02 per cent per degree centigrade over a range of $-55°$ to 70°C. Batteries have a thermal time constant of 15 to 30 min depending on size, construction, and material.

The most important requirement for constant potential of dry batteries is the complete absence of short-circuiting or even of appreciable current drains. If short-circuited briefly, a battery will recover to a steady potential of somewhat lower value, but its life as a constant potential element is materially reduced.

A series of tests[1] on Burgess 5156 batteries showed them to be constant to better than 0.05 per cent over a period of 1 to 3 months. Current drain was less than 0.1 μa except during periodic measurements when the current went up to a few microamperes. During these tests, the tem-

[1] H. Sack, *Constancy of EMF's of Dry Batteries*, Cornell University, Oct. 2, 1945.

perature ranged from 20° to 25°C. An electronically regulated power supply using several Burgess Z30N batteries as a reference when checked over a period of 20 days was found to be constant to about 0.05 per cent.

Because of the large potentials available from dry batteries a comparatively insensitive comparison circuit can often be used with a dry battery reference element.

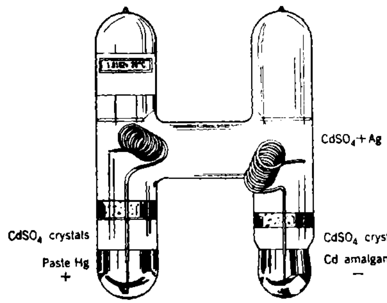


FIG. 15·1.—Standard cell.

Standard Cells.[1]—In applications where a more precise voltage reference is required than that provided by dry batteries, unsaturated cadmium cells (Weston type) are of value. These cells (Fig. 15·1) have a potential of slightly over 1 volt which is constant within ±0.1 per cent of the original voltage over a temperature range of −16° to 50°C. This cell can be tilted to 110° without introducing errors; but when inverted, the cell is open-circuited. Currents of the order of 100 μa for a short time do not cause permanent damage.

The most serious difficulty in the use of standard cells as reference elements is the requirement of better than ±100-μv stability of the equating circuit.

Glow Tubes.—Glow tubes of the neon-bulb and VR-tube variety are very popular regulator voltage reference elements in electronic instruments.

Glow tubes are very nonlinear resistance elements; for a considerable current range, the voltage across the tube is nearly constant. The construction of glow tubes is very simple. Two electrodes of iron, nickel, aluminum, or magnesium in a low-pressure gas or mixture of gases (usually hydrogen, nitrogen, helium, neon, or argon) make up a glow tube. By choice of gas mixture and electrode geometry a range of 50 to several thousand volts operating potential is obtainable. The ordinary commercially available tubes (Fig. 15·2) have operating potentials of 55 to 150 volts. Glow discharge tubes require a potential 25 to 50 per cent higher than the operating potential to initiate the glow discharge and require a minimum current to maintain the glow discharge. These tubes generally function in a satisfactory manner over a wide temperature range, in excess of −65° to +80°C. Vibration and change in position do not adversely affect the stability of these tubes. The striking potential is in some cases affected by the photoelectric proper-

[1] "Eppley Standard Cells," *Bull.* 1, The Eppley Laboratory, Inc., Newport, R.I., July 1941.

ties of the electrode material. Table 15·1 is a summary of the characteristics of neon bulbs and VR tubes.

Neon bulbs vary considerably in characteristics from tube to tube, whereas VR-tube characteristics are constant within a few per cent due to more careful construction and more thorough testing. The useful current range as given in Table 15·1 is conservative and, in general, is the range over which the tube should be used as a voltage reference. Currents between 5 and 40 ma are acceptable for the VR-75, VR-105,

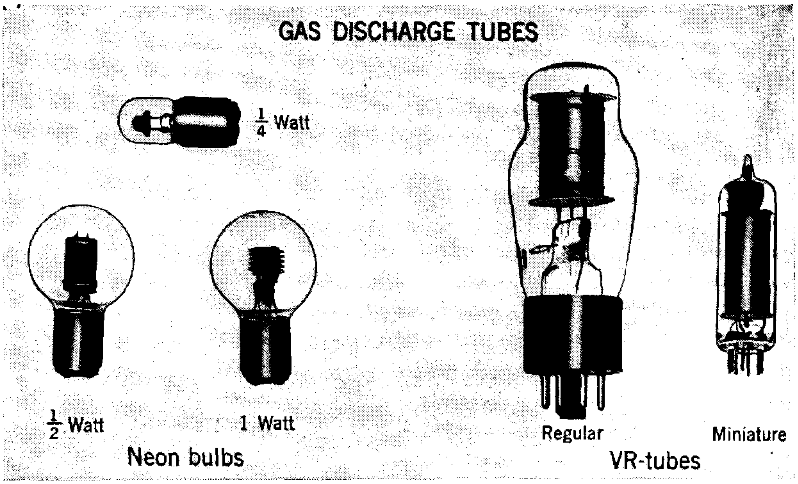


FIG. 15·2.—Gas discharge reference elements.

and VR-150 tubes. Higher currents shorten tube life, and lower currents result in erratic operation.

The desirable characteristics of glow tubes as voltage reference elements differ somewhat from those for simple regulators. Effects of time and temperature on the voltage at a constant current are of primary importance in applications of glow tubes as reference elements. The exact operating voltage, if large, is of relatively little importance in the choice of a glow tube for a voltage reference. Wide range of current and low dynamic impedance are the important features of glow tubes for simple regulators.

The volt-ampere characteristic, of prime importance in simple regulator elements, is of secondary importance for reference elements. Although a small slope is desirable, one may compensate for greater slope by more gain in the amplifier to obtain the desired degree of regulation. Figure 15·3 shows the modes in a typical volt-ampere curve for a VR-105.

TABLE 15·1.—SUMMARY OF GASEOUS DISCHARGE TUBES

Type	Size	No. tested	Average voltage at current	Useful current range, ma	% increase over useful current range	Minimum starting voltage	Temp. coef. for average, ppm/°C	Current for drift test, ma	Average drift rate for average tube at 15 hr, mv/min	Notes
Neon NE2	0.04 w	1	58 v 0.2 ma	0.2–1	8.00	75	......	0.1	50	Characteristics quite variable
Neon NE16	$\frac{1}{4}$ w	1	52 v 0.5 ma	0.5–10	18.00	85	......	..	....	Characteristics quite variable
Neon NE29	$\frac{1}{2}$ w	1	95 v 1 ma	1–10	15.00	125	......	..	....	Characteristics quite variable
Neon NE32	1 w	1	58 v 1 ma	1–20	22.00	75	......	..	....	Characteristics quite variable
991	....	..	55 v 0.5 ma	0.5–2	.....	87	......	..	....	Same size as $\frac{1}{4}$ w neon A8-67 v. Voltage limits for 0.5 to 2 ma
OA3 VR75	Regular octal	8	73 v 10 ma	10–30	3.17	105	−50	15 30	0.12 0.20	....
OC3 VR105	Regular octal	34	105 v 10 ma	10–30	0.70	133	+30 −50	15	0.20	....
OD3 VR150	Regular octal	9	150 v 10 ma	10–30	0.40	185	−200 +70	15	0.20	....
VR105 S860	Miniature	10	107 v 10 ma	10–30	0.67	135	+20 −60	15	0.25	....

TABLE 15·1.—SUMMARY OF GASEOUS DISCHARGE TUBES.—(*Continued*)

Type	Size	No. tested	Average voltage at current	Useful current range, ma	% increase over useful current range	Minimum starting voltage	Temp. coef. for average, ppm/°C	Current for drift test, ma	Average drift rate for average tube at 15 hr, mv/min	Notes
VR105 HD-52	Miniature	10	105 v 10 ma	10–30	1.52	135	+200 −60	15	0.20	....
VR105 S901	Miniature	3	105 v 2 ma	2–4	4.00	135	+400 −400	3	0.10	....
VR150 OA2 (S856)	Miniature	6	151.5 v 10 ma	10–30	6.67	185	+300 −300	15	0.40	....
VR 150 OA2	Miniature	10	148	10–28	0.33	185	$\begin{Bmatrix}+75\\-50\end{Bmatrix}$	..	....	....
VR 150 HD 51	Miniature	10	148	10–27	1.00	185	$\begin{Bmatrix}+70\\-40\end{Bmatrix}$	15	0.42	....
VR 75 CV 284	British miniature	1	73	10–20	3.67	...	−350	6	0.33	....
VR 95 CV 286	British miniature	1	92	5–10	4.82	...	$\begin{Bmatrix}+80\\-60\end{Bmatrix}$	6	0.60	....

Figure 15·4 (scope picture redrawn) shows the effect of increasing current (lower trace left-hand side) and decreasing current (upper trace) on the modes of operation.

The volt-ampere curve (Fig. 15·3) was chosen as an example of the characteristics of a glow tube with very distinct modes in order to show more clearly the effects of changing from one mode to another. The voltage difference between two modes for a VR tube is of the order of 0.1 to 0.5 volt. This change in mode of operation occurs when the tube is ignited, when the current is changed, or for other reasons which are not always apparent. Mode changing is primarily due to time variations in the spatial distribution of the glow discharge on the electrode system and to variations in the current.

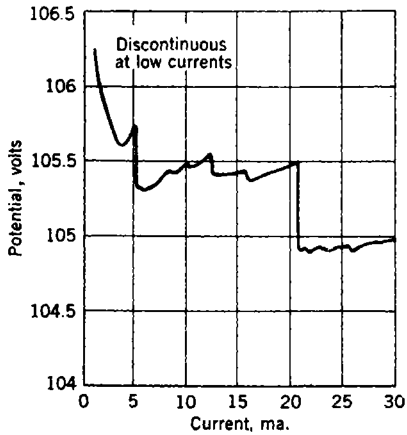


FIG. 15·3.—OC-3 (VR-105) glow tube volt-ampere characteristics.

Figure 15·5 shows the effect of mode changes over a period of time during which the tube was turned off every 4 min for 20 sec. An unusual characteristic of this effect is the very low frequency of the occurrence of large changes, although these changes may amount to $\frac{1}{2}$ volt for a 105-volt tube. A statistical analysis of considerable data on VR-105's indicates that the probability of any randomly chosen instantaneous voltage differing from the mean by more than 0.1 volt is less than 10 per cent.

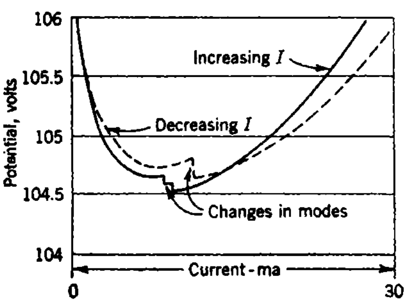


FIG. 15·4.—Volt-ampere characteristics of VR-105 for increasing and decreasing current.

Temperature effects on glow tubes are of two types: a reversible change of the voltage with temperature and an irreversible change. The former is produced by reducing the temperature of the glow tube from room temperature, while the latter occurs when the glow tube is heated to 80°C. The time-voltage curve of Fig. 15·5 shows such an irreversible temperature effect. An interesting feature of the VR-105 and many other tubes is the fact that the temperature coefficients are of opposite signs for deviations above and below room temperature. A small number of tubes tested had variations of less than $\frac{1}{2}$ per cent over the temperature range from 25° to −50° and less than $\frac{1}{2}$ per cent from 25° to 80°C.

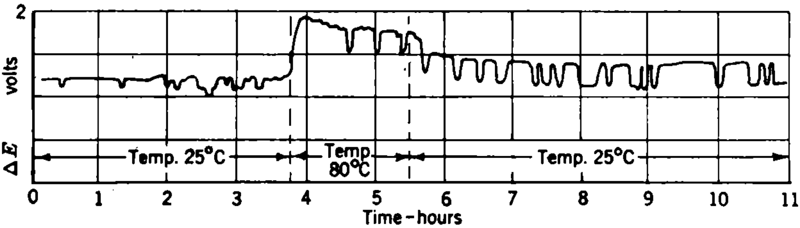


Fig. 15·5.—Time, voltage, and temperature characteristics of a glow tube. Tube switched off for 20 sec every 4 min.

Since temperature variation does not affect the current at which changes in modes occur, the appropriate current for a VR tube as a reference element can fortunately be chosen to minimize to a great extent the possibility of error due to changes in mode of operation. VR-105 tubes have been found to have maximum stability when operated with a steady current of 12 ma.

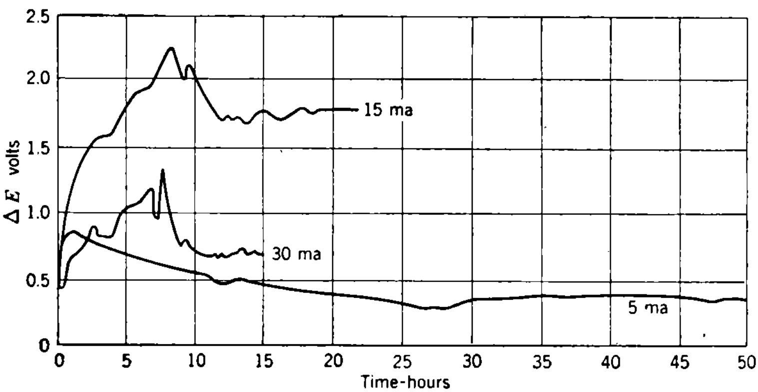


Fig. 15·6.—Drift data, for RCA OA-2 (VR-150).

Drift of voltage with time is of considerable importance in the use of glow tubes for reference elements. The effect of age upon the voltage of a glow tube is shown in Fig. 15·6. The drift is less for lower currents for an average tube. These typical characteristics show a large initial drift followed by a rather steady drift of 200 μv/min continuing for the operating life of the tube. From all available information the drift appears to be statistical with a tendency for positive drift for the first part of the tube life.

Constant-current Elements.—Any one of a number of filaments of various metals may be operated at elevated temperatures as a nonlinear element that may be used in series with a load as a current regulator or in a bridge or divider network as a reference element. The filament is usually placed in a vacuum or in an inert gas atmosphere to reduce oxidation and to realize more nearly the desired thermal properties.

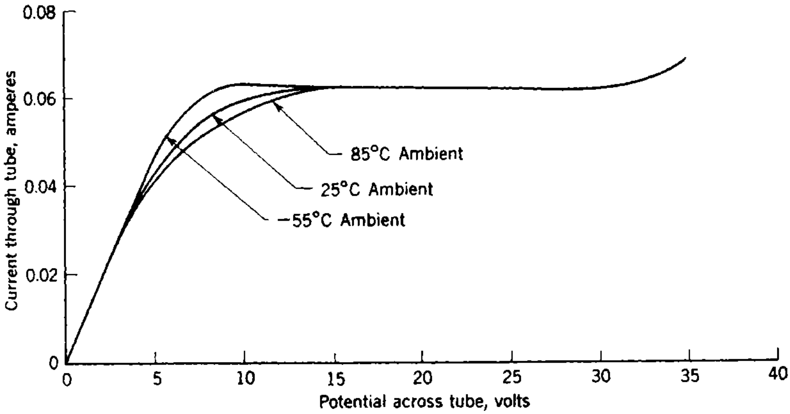


FIG. 15·7.—Volt-ampere characteristics of Amperite ballast tube.

There are two classes of filamentary elements. One class, consisting of an iron filament in an atmosphere of hydrogen, provides good current regulation, but its characteristics vary considerably with ambient temperature. These elements depend on the heating of the filament for change in resistance and have a time constant, depending on construction, of a fraction of a second to several seconds. Typical volt-ampere characteristics for an Amperite current regulator (ballast) are given in Fig. 15·7. The approximate curve for any other Amperite can be obtained by appropriate change of current or voltage scale. Amperite tubes are commercially available to regulate at current values from 0.050 to 8 amp and have threshold values from 0.4 to 60 volts. After being aged for approximately 12 hr the tubes show no further appreciable changes in characteristics.

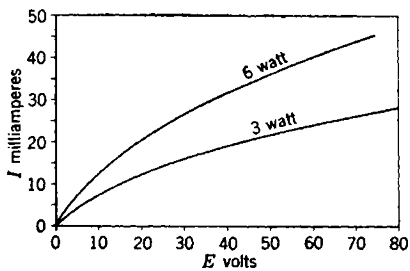


FIG. 15·8.—Volt-ampere characteristic of 110-volt Mazda lamps.

The other class of filamentary elements has a high-temperature tungsten filament in a vacuum, resulting in a reduction of the effect of ambient temperature on the volt-ampere characteristics. This class of filamentary elements is often used in resistance-capacitance oscillators to stabilize the output amplitude, an ordinary 3- or 6-watt 110-volt Mazda lamp being used in a divider to adjust the amount of negative feedback auto-

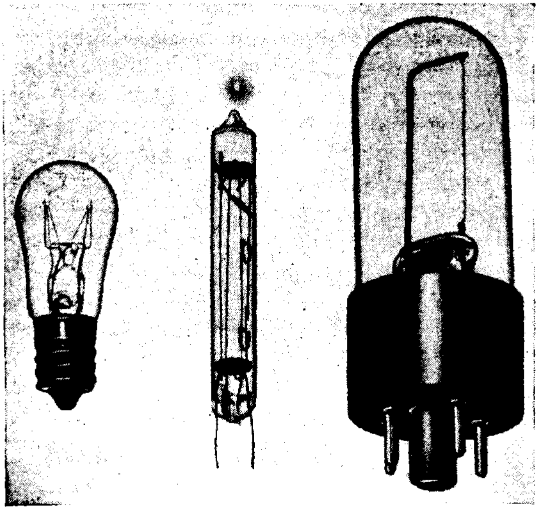


FIG. 15·9.—Filamentary reference elements.

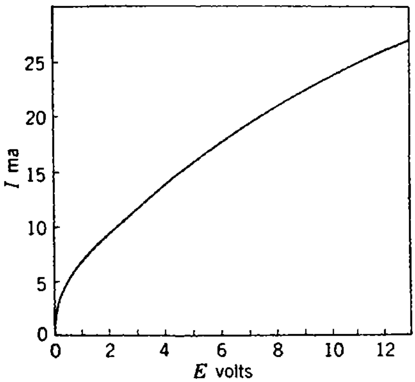


FIG. 15·10.—Volt-ampere characteristics of Victoreen filament-type reference tube.

matically in a majority of the circuits. Volt-ampere characteristics for these two Mazda lamps are shown in Fig. 15·8.

An experimental tube was developed by the Victoreen Instrument Co. (Fig. 15.9, center) in an effort to obtain an efficient stable reference element for regulated d-c supplies that must be better than $\frac{1}{10}$ per cent. The volt-ampere characteristics are given in Fig. 15·10. When operated at 10 volts and 24 ma in a suitable resistance network across a 250-volt d-c supply, 53-mv output is realized for each per cent change in the supply. For an output accuracy of $\frac{1}{10}$ per cent a comparing circuit used with such an element would have to have stability of better than 5 mv. A number of the resistance elements may be connected in series for a larger output.

Miscellaneous Nonlinear Impedances.—There are two broad classes of nonlinear element: those which are nonlinear resistances due to other than thermal effects, such as Thyrite and glow tubes, and those which are

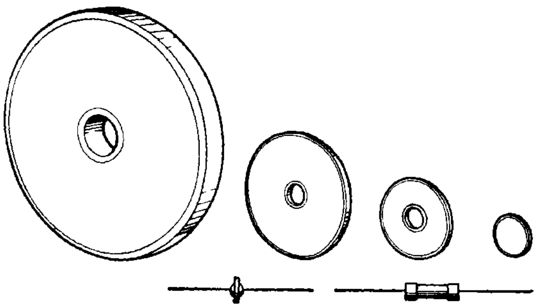


FIG. 15·11.—Typical Thyrite units.

nonlinear due to thermal effects, such as Thermistors, diodes, and filamentary elements.

In the use of nonlinear reference elements, the volt-ampere characteristic is very important because it shows how effectively the element may be applied to a particular circuit. Other important characteristics are current requirement, temperature coefficient, temperature limits, change of volt-ampere characteristics with age, voltage rating, effects of vibration, size, weight, and frequency response.

Thyrite.—Thyrite is the trade name of a nonlinear resistance material produced by the General Electric Company. Figure 15·11 shows various commercial units.[1] This material is made of silicon carbide with a ceramic binder. A metal coating is sprayed on the surfaces to provide electrical contact. The instantaneous and steady-state volt-ampere characteristic of Thyrite is given by the equation

$$I = KE^n. \tag{1}$$

The quantity K depends on the resistivity and dimensions of the particular Thyrite unit. The exponent n is at least 3.5 and can be as high as 7 in special cases. Figure 15·12 shows volt-ampere characteristics of a number of commercially available types.

The volt-ampere characteristics are the same for impulses of microseconds duration as for direct current and there are negligible polarizing

[1] From *GE Thyrite Bull.* GEA 4138. Western Electric Co. *Varistors* are similar in composition and characteristics.

effects. Voltage and current are almost exactly in phase, and the volt-ampere characteristic is symmetrical for both positive and negative polarity.

A continuous rating of $\frac{1}{4}$ watt per square inch is allowable in still air. A short time temperature rise of 80°C results from an input of 2000 watt-sec per cubic inch.

The most serious limitation to the use of Thyrite for reference elements is its high temperature coefficient. At constant voltage, current

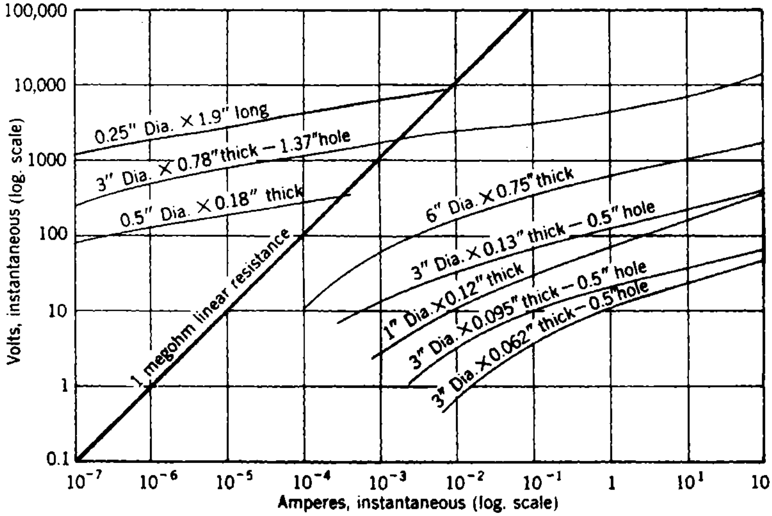


FIG. 15·12.—Volt-ampere characteristics of Thyrite.

increases 1 per cent per degree centigrade. Thyrite may operate at 110°C continuously. Humidity has very little effect on properly impregnated units.[1]

Thermistors.—Thermistors[2] are nonlinear elements that depend on power dissipated in the element to raise the temperature and decrease the resistance. There are many types with widely varied characteristics. Thermistors are made by Western Electric of a mixture of metallic oxides. While there are many possible combinations of materials, nickel oxide appears in most mixtures with manganese, uranium or

[1] Considerable information on circuit applications may be found in T. Brownlee, "The Calculations of Circuits Using Thyrite," *Gen. Elec. Rev.*, April 1934, pp. 175–179, and May 1934, pp. 218–23.

[2] J. A. Becker, C. B. Green, and G. L. Pearson, "Properties and Uses of Thermistors—Thermally Sensitive Resistors," *Elec. Eng.*, **65**, Trans. 711–725, November 1946.

copper oxides, and other materials for binders. The very small bead units are made in a number of mechanical variations of which types D-163903 and D-166382 are two examples. Curves for typical units are

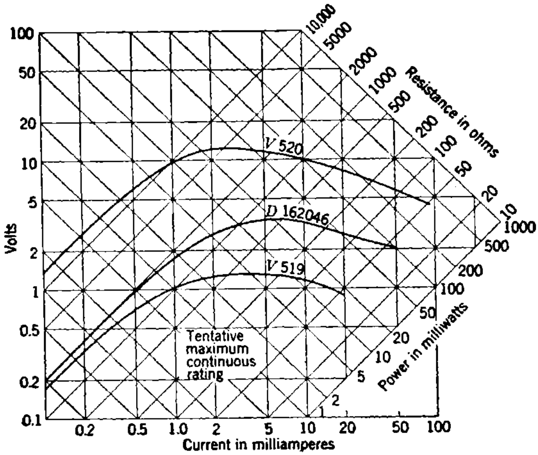


FIG. 15·13.—Typical Thermistor characteristics at 25°C in still air.

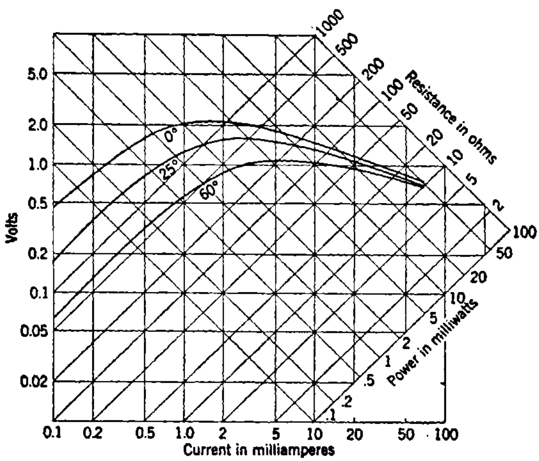


FIG. 15·14.—Variations in Thermistor characteristics with ambient temperature. These data are for D-163903 and D-166382 units.

shown in Fig. 15·13. Figure 15·14 indicates the variations in characteristics with ambient temperature of type D-166382.

The variation of the resistance of a Thermistor with ambient temperature and power dissipation may be indicated by the following formulas:

$$\left.\begin{aligned} R &= R_0 e^{(B/T)-(B/T_0)}, \text{ and} \\ W &= C(T - T_0), \end{aligned}\right\} \qquad (2)$$

where R_0 = resistance, ohms, at temperature T_0, degrees Kelvin,
B, C = constants, and
W = power dissipation.

The thermal time lags vary from a small fraction of a second with the small bead type D-166382 to several hours for large units.

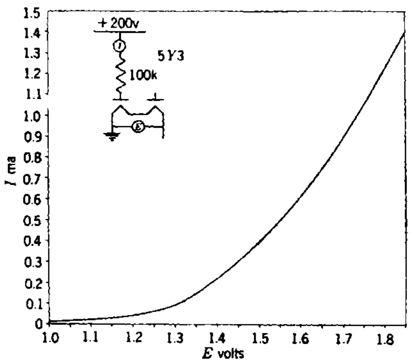


FIG. 15·15.—Typical plate current vs. filament voltage characteristics of 5Y3 as emission limited reference element.

The ratio of voltage output to change of voltage input of the bridge or divider sampling circuit utilizing Thermistors of course depends greatly on the circuit and component details. In a specific case where all factors were carefully chosen the ratio of percentage output change to percentage input change was 4:1.

Thermistors, like nearly all nonlinear elements, cannot be manufactured to very close specifications. A typical Thermistor D-166382 has tolerances at a particular temperature of 25 per cent for resistance at no current and at 25 ma. The power level tolerance at 250 ohms is also as high as ± 25 per cent. However, as the result of considerable development effort by Bell Telephone Laboratories, Thermistors are very stable in their characteristics even though characteristics may vary from unit to unit.

Thermionic Vacuum Tubes as Reference Elements.—A diode may be used as a reference element in a regulator circuit. Figure 15·15 shows the plate-current change for a small percentage filament-voltage change. A plate current of about 100 μa and a filament voltage of 1.33 volts are suitable operating conditions. Ten tubes were checked, and filament voltages for plate currents of 100 μa were fairly evenly distributed over a range of 1.20 to 1.35 volts.

As a reference element a diode should be operated in a bridge circuit such as is illustrated in Fig. 15·16. To obtain the greatest ratio between a-c voltage into the diode and the voltage out to the control tube grid a high resistance should be associated with the diode and a high voltage should be used across the bridge. The time constant of this circuit depends on the thermal lag of the filament; it may be a fraction of a second to several seconds depending on construction. Such oxide-coated cathode thermionic elements operated at heater voltages below manufacturers' ratings are apt to have short life unless current is kept very low. The explanation for this lies in the fact that at low heater temperatures the cathode coating has a tendency to be destroyed at the surface faster than replacement molecules of the oxide can diffuse to the surface. Any such low-voltage application should therefore be subject to careful life tests. As with other nonstandard characteristics of vacuum tubes, such tests should be made on representative samples of tubes made by all or at least many manufacturers.

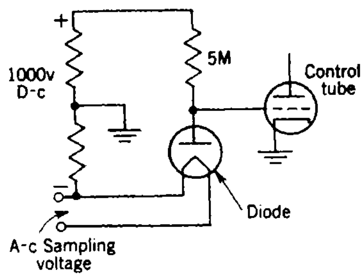


FIG. 15·16.—Bridge circuit for diode emission reference element.

15·3. Sampling Circuits.—The function of the sampling circuit is to provide a fraction of the regulator output or to provide some other voltage for comparison with the output of the reference element in order to derive an error signal that may be used to operate the control element. In some regulator systems, the functions of the sampling circuit, comparing circuit, and reference are all performed by one nonlinear network, and any division of this type of circuit into the above circuits is artificial. Nevertheless, there are a sufficient number of circuits that divide logically into these subcircuits to make the system of classification a useful one.

In typical regulators, the regulator output is varied in such a way as to make the sampling circuit output equal to the reference element output. Consequently, any variation in the percentage of the regulator output that is taken as a sample is reflected directly in the regulator output. Since most sampling circuits take only a portion of the output for comparison with the reference element, small variations in the sampling circuit may lead to rather large (absolute) variations in the output.

There are three broad classes of sampling networks: linear networks, nonlinear networks, and compensating circuits. The term "compensating circuits" is here defined as a network introducing into the sampling circuit a voltage proportional to or at least a direct function of the input voltage and/or load current in such a way as to increase the regulation factor and/or reduce the source impedance. Many regulator circuits employ more than one class; for instance, compensating circuits may be used in conjunction with linear networks to reduce the output impedance of the regulators. In general, linear networks have the best static stability. Nonlinear networks are of value because they have less d-c attenuation than linear networks; hence, for some applications, the increase in gain may more than make up for the poorer stability. Compensating networks are very useful when regulation is required over a large range of input voltage; by proper sampling of the input voltage the requirements on the regulator output sampling circuit are reduced.

Linear Networks.—Voltage dividers made up of resistors are the usual sampling circuits. Figure 15·17 shows two examples of this type of

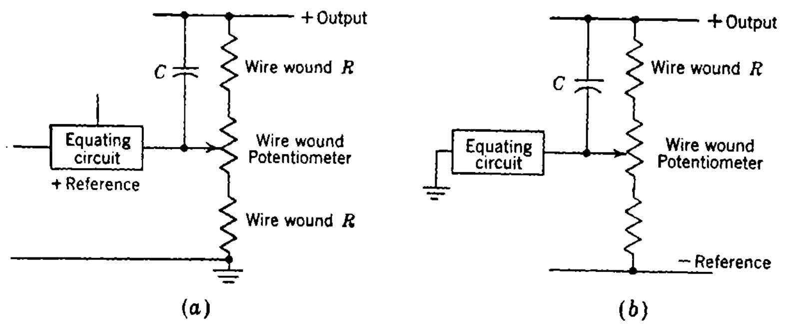


FIG. 15·17.—Representative sampling circuits.

circuit. For maximum stability it is necessary to use wire-wound resistors, both for the fixed and for the adjustable components. If a potentiometer is used, it should be no larger than is necessary to provide the required range of control; this precaution minimizes the effect of the discrete values of resistance obtainable from wire-wound potentiometers and the effects of the temperature coefficient of the potentiometer.

In practically every case, sampling circuits cause a loss of gain around the degenerative loop, but this is normally not a serious matter. Since most comparison circuits require the outputs from the sampling circuit and the reference element to be at approximately the same voltage level, higher reference voltages mean smaller sampling circuit gain reductions. The loop gain at audio frequencies can be increased by capacitively coupling directly from the output of the regulator to the input of the comparing circuit. Such an arrangement does not affect the d-c attenuation but does greatly reduce the a-c attenuation. This reduces the output ripple at these frequencies and in addition contributes to stability of the loop if correctly designed.

For certain applications, particularly for low-voltage, high-current regulation a solenoid or induction device is useful as a sampling circuit. For these applications a spring may be used as a reference element.

Nonlinear Networks.—Figure 15·18 shows two applications of nonlinear networks as sampling circuits. In circuits such as these it is meaningless to separate the sampling circuit from the reference element.

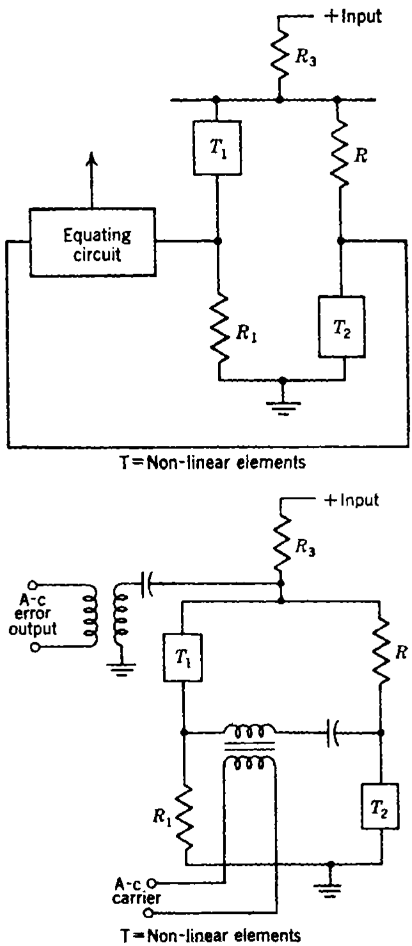


FIG. 15·18.—Sampling circuits with nonlinear elements.

Compensating Networks.—In many degenerative regulators the regulation factor and the output impedance may be improved by the use of suitable compensating networks associated with the sampling circuit.

A compensation circuit has been defined above as a circuit that introduces voltages proportional to input voltage and/or load current into the sampling circuit in such a way as to increase the regulation factor and/or reduce the source impedance.

Compensation is most practical in electronic regulators that have relatively low loop gain and are operated over a limited range of line and load variations. Linearity of sampling, comparison, amplifier, and control circuits determines the extent to which compensation can improve the operation of a regulator circuit.[1]

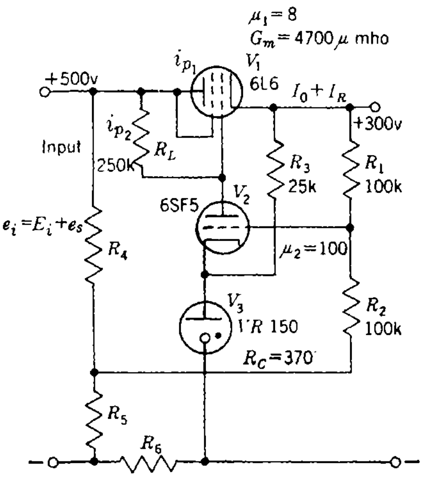


FIG. 15·19.—Compensated regulator.

Circuits, analysis of operation, and results of use of compensation of a regulator for line and load variations have been described in the literature.

The following equations of conditions to make the regulation factor infinity and the source impedance zero have been taken from Hill's article.

To make the regulation factor high with the circuit of Fig. 15·19 the following relation between R_4 and R_5 must be realized:

$$\frac{R_5}{R_4 + R_5} = \frac{\left(1 + \frac{1}{\mu_1}\right)\left(1 + \frac{\mu_2 R_c}{r_{p_2} + R_L}\right)}{\left(\frac{R_1}{R_1 + R_2}\right)\left(\frac{\mu_2 R_L}{r_{p_2} + R_L}\right)} - \frac{R_1 - R_2}{\mu_2 R_1}. \tag{3}$$

If R_4 is chosen to be 100 k and with other values as shown in Fig.

[1] F. L. Hogg, "Regulated Power Supplies," *Wireless World*, November and December 1943; W. R. Hill, Jr., "Analysis of Voltage-Regulator Operation," *Proc. IRE*, **33**, 38–45, January 1945; F. V. Hunt and R. W. Hickman, *Rev. Sci. Inst.*, **10**, 6 (1939).

15·19, an optimum value of R_5 of 1,160 ohms was found experimentally, in reasonably good agreement with the calculated value.

With compensation, the regulation factor of this circuit was improved by more than a factor of 10 over an input range of 25 per cent with a 40-ma load current.

The use of a compensating circuit to reduce the effective source impedance of a regulator to nearly zero by a series resistor R_6 in the sampling circuit is also shown in Fig. 15·19.

The value of R_6 is low, usually less than 10 ohms, and may be calculated by the equation[1]

$$R_6 = \frac{1 + \dfrac{\mu_2 R_c}{r_{p_2} + R_L}}{g_{m_1} \dfrac{\mu_2 R_L}{r_{p_2} + R_L} \dfrac{R_1}{R_1 + R_2}}. \tag{4}$$

The computed value of R_6 for the values given was 6.05 ohms, and the experimental value was 7.0 ohms at a load of 40 ma. At an input of 500 volts and a current range of 0 to 100 ma the source impedance was negative over part of the range and the value was of the order of 2 ohms. The source impedance without compensation was about 6 ohms; the improvement of a factor of 3 is probably not worth while in most instances because R_6 would have to be adjustable and there is danger of oscillations occurring.

The main disadvantages of compensation methods are that circuit adjustment may be required when tubes are replaced and, in the case of the lowered source impedance circuit, a "floating" supply is required.

15·4. Comparison Circuits.[2]—The comparison circuit in a degenerative regulator produces a signal that is a measure of the magnitude and sense of the difference in potential between the sampling circuit and the reference element. There are two broad classes of comparing circuits: direct-coupled amplifiers and modulators. The direct-coupled amplifiers have two inputs, one for the output of the reference element and one for the output of the sampling circuit; the output of the direct-coupled amplifier is the amplified d-c difference between these quantities. A modulator performs the same discrimination function except that the output is an a-c error voltage whose amplitude is at least roughly proportional to the magnitude of the error and whose phase indicates the sense of the error.

The most important characteristics of comparison circuits are the dynamic response and static stability, sensitivity, and effects on

[1] This corrects an error in Hill's paper.

[2] See Vol. 18 for detailed treatment of d-c amplifiers.

the reference element and sampling circuit. In d-c amplifier circuits the sensitivity is usually much better than the static stability; hence the static stability determines the utility of the circuit. Static stability is measured quantitatively by the change in grid-cathode bias necessary to maintain constant plate current as the tube ages or as heater voltage changes. Static stability is usually expressed, somewhat arbitrarily, as change in bias (ΔE_c) per unit time or for a 20 per cent change in heater voltage. For tubes used in differential-amplifier circuits the differential change in bias is more important than the change in bias of either tube. Direct-coupled amplifier comparing circuits have good dynamic response. Modulator comparing circuits, while having excellent static stability, are limited in their dynamic response, since the error information is available only at the carrier frequency.

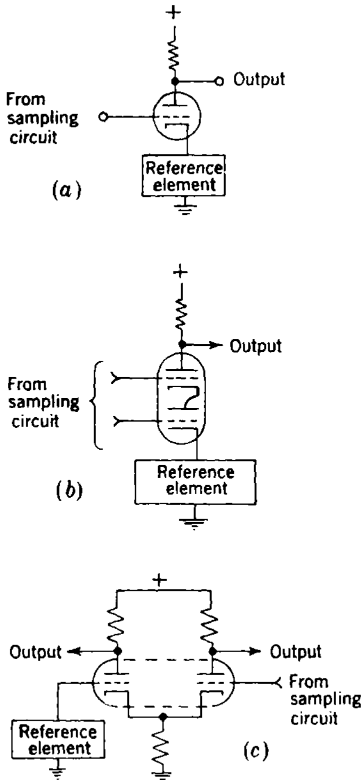


FIG. 15·20.—Direct-coupled amplifier comparison circuits. (a) Single-ended amplifier; (b) cascode amplifier; (c) balanced amplifier.

Direct-coupled Amplifiers.—Figure 15·20 is a schematic of the three principle types of direct-coupled amplifier circuits used as comparing circuits. The amplified output of each of these circuits is a measure of the difference voltage between the sampling circuit and the reference element.

The single-ended amplifier (Fig. 15·20*a*) usually employs a multigrid tube to provide high gain. The single-ended amplifier has the characteristic of passing current through the reference element, an undesirable feature with many reference elements. A further disadvantage is that no compensation is provided for changes in cathode-grid voltage due to heater voltage fluctuations and aging.

The cascode amplifier (Fig. 15·20*b*) has the same inherent disadvantages as the single-end amplifier but has the advantage of very much higher gain. The increased gain reduces the effective output impedance and hence improves the regulation. The static stability is limited by the ΔE_c characteristics of only the lower tube.

The balanced amplifier (Fig. 15·20c) loads neither the reference element nor sampling circuit and balances out a large percentage of the effects of changing heater voltage and aging. The use of a dual tube is hence recommended where possible.

Table 15·2 is a summary of the results of tests on a number of tubes to illustrate the effects of heater voltage variations for a comparison tube. The desirability of keeping the plate current low is shown by the lower values of ΔE_c and $\Delta\Delta E_c$ for lower currents.

TABLE 15·2.—SUMMARY OF TUBES FOR EQUATING CIRCUITS

Type of tube	No. tested	Single or balanced	Plate current I_p, ma	ΔE_c for 20% change in E_h, mv			$\Delta\Delta E_c$* for 20% change in E_h, mv		
				Min.	Max.	Average	90% under	Max.	Average
6SJ7	10	S	2.5	182	308	216	...	...	..
		S	10.0	212	936	410	...	...	..
6AC7	10	S	2.0	180	336	210	...	...	..
		S	8.0	204	700	310	...	...	..
6AG7 (RCA)	10	S	10.0	197	375	230	...	...	..
		S	40.0	231	663	317	...	...	..
6SN7	10 (20 sections)	Both	2.0	183	313	210	30	40	14
		Both	10.9	228	624	470	150	180	83
6V6	....	S	56.0	111	677	400	...	...	..
6SL7	18 (36 sections)	B	0.1	189	228	...	20	...	..
6SU7 Tung sol	67	B	0.1	...	...	...	25	...	..

NOTE: All groups of tubes are of assorted makes, except as noted.
* Difference between sections of a twin tube in balanced amplifier.

Fluctuating effects have about the same magnitude for diodes, triodes, tetrodes, and pentodes; for high and low μ's; and for high and low transconductances. In fact, these effects appear to be common to all oxide-coated unipotential cathode structures. At or below 25 per cent of the allowable average current, the fluctuating voltage is about 0.25 volt (measured at the grid or cathode) for a 20 per cent change in heater current. Aging effects are similar to the effects of changing heater potential. From the standpoint of this fluctuating potential the 6SL7 is as low as or lower than any other common tube. Since it is a dual tube, the differential fluctuation voltage is quite low. Tests indicate that after proper aging, 90 per cent of 6SL7's or 6SU7's operated at low plate currents—0.1 to 0.4 ma—have differential errors under ± 25 mv for a 20 per cent change in heater voltage and under errors ± 25 mv differential over a period of one week at constant heater potential.

Permanent changes of E_c of the order of 10 to 20 mv may result from mechanical shocks which cause permanent changes in the tube geometry. This effect is small compared with other factors in single-ended circuits but may be important in balanced circuits, as these other factors tend to cancel out and one tube may go plus and the other go minus; hence $\Delta\Delta E_c$ may be 20 to 40 mv.

Change in E_c with time, even when E_h is constant, is due primarily to changes in the oxide-coated cathode. This effect is quite serious and in a period of 7 days may amount to as large a change in E_c as would be produced by a 10 per cent change in E_h.

Drift tests on five pairs of 6SJ7's, connected in a differential d-c amplifier circuit operating at about 0.1 ma, constant E_h, and at about 100 volts between plate and cathode, showed an average differential drift rate of about 7 μv/min for 40 to 60 hr. Thus, for a period of 50 hr, differential drift of less than ± 50 mv should be realized.

Modulator Equating Circuits.—There are three classes of modulator discriminator circuits. Triodes (or multigrid tubes) depend on curvature of tube characteristics; diodes depend on discontinuity between two (nearly) linear characteristics; and moving contact vibrators and breakers depend on switch action. Each produces, when actuated by a carrier, an a-c signal that indicates by its phase and amplitude the voltage difference between the sampling circuit and reference voltage.

Tube modulators have essentially the same limitations of stability as have d-c amplifiers, as nearly all instability is due to the characteristics of the oxide-coated cathode.

Three balanced-modulator circuits are shown in Fig. 15·21; two use double diodes, one a double triode. These circuits have two cathodes each which tend to balance out E_c fluctuations, giving about the same stability as direct-coupled amplifiers. Figure 15·22 shows a typical circuit using a modulator to obtain a d-c voltage that is an amplified measure of the voltage $E_2 - E_1$ and may be used to operate a control element.

Mechanical switch modulators used as comparison circuits provide a much higher order of stability than electronic tubes. Under favorable conditions it is possible to detect a voltage difference of the order of 10^{-9} volt at 5 ohms impedance level with the General Motors Research Laboratory modulator circuit.[1] For regulator applications, because of the higher impedances involved, accuracies of the order of 1 mv are usually realized with switch type modulators.

The Brown Converter vibrator and the Leeds and Northrup vibrator

[1] W. P. Wills, "D-c to A-c Conversion Systems," *Elec. Eng.*, **66**, 39–40, January 1947; M. D. Liston, C. E. Quinn, W. E. Sargeant and G. G. Scott, "Contact Modulated Amplifier to Replace Sensitive Suspension Galvanometers," *Rev. Sci. Inst.*, **17**, 194–198, May 1946.

are examples of single-pole two-position switches which may be used in comparing circuits similar to Fig. 15·22. These units (Fig. 15·23*a* and *b*), are especially designed to obtain a high order of accuracy and stability.

The Brown Converter consists of a vibrating reed which forms the center contact of a single-pole two-position switch. The reed is driven electromagnetically and makes intermittent connection to two contacts The complete unit is encased in a metal shell to provide mechanical, electrical, and dust shielding and is designed to plug into a standard tube socket. The commercially available unit has a reed with a resonant frequency of approximately 90 cycles and is designed to operate at 50 to 60 cycles, 6.3 volts. The outside contacts are normally shorted together through the reed for 7 per cent of the operating cycle, although they may be adjusted for other types of operating conditions. Figure 15·22 illustrates normal contact shorting. The unit is insensitive to temperature and mounting position. Its life is long when required to break no

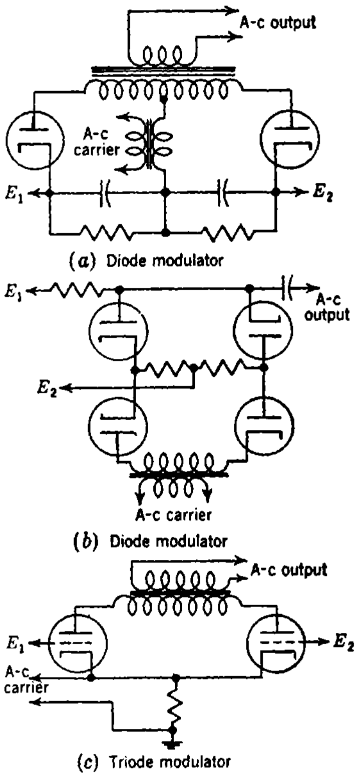


FIG. 15·21.—Vacuum-tube modulators.

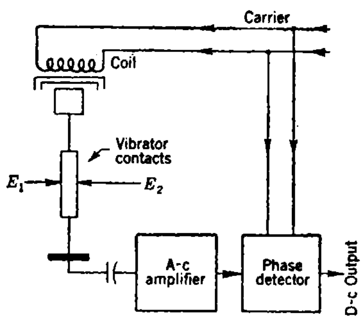


FIG. 15·22.—Switch-type modulator comparison circuit.

more than 10 volts, but contact life is materially shortened if operated at more than 16 to 18 volts.

The limit of the sensitivity of the Brown Converter is determined by the impedance of the circuit in which it is used. Even with considerable shielding there is capacitive coupling between the coil and the reed of the order of 5 $\mu\mu$f; this gives, for instance, 1 mv of "noise" for an input impedance of 100,000 ohms.

Switch-type modulators give the maximum static stability of all comparing circuit designs; but as with all modulator comparing circuits, they are limited in dynamic response by the rate at which the information is gathered. For the Brown Converter this limit is 60 cycles. Vibartor modulators have been built to operate at higher carrier frequencies, but they are still experimental.

Miscellaneous Devices.—Another modulator[1] of particular interest because of its high frequency and accuracy is based on the elements of a

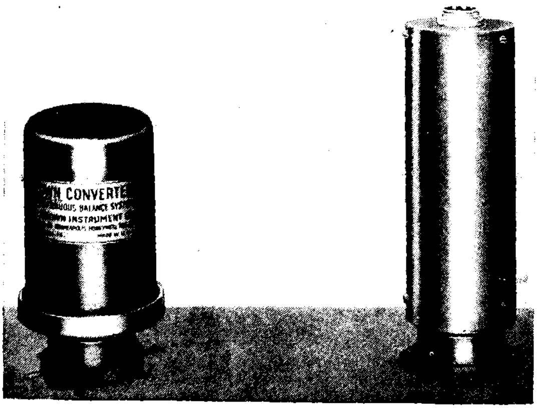


FIG. 15·23.—Vibrator equating devices.

condenser microphone and has been used by RCA and others for equating and as a linear modulator in computers. Figure 15·24 shows the RCA unit disassembled, and Fig. 15·25 shows the use of this unit in a modulating circuit. When d-c potentials are applied to the plates through high impedances, the change in capacity between the plates and the diaphragm causes an a-c potential to appear at the plates. The amplitude of the alternating current is proportional to the d-c potential difference between the diaphragm and the plate and the amplitude of the diaphragm vibration. The diaphragm is driven by the "voice coil," and the amplitude is controlled (the amplitude can be constant or controlled and variable) by a feedback amplifier with one of the plates at fixed or variable potential as a reference.

In the circuit shown in Fig. 15·25 the diaphragm is designed to be operated at ground potential. The diaphragm is mechanically resonant

[1] See, for example, S. Rosenfeld and W. M. Hoskins, "Modified Zisman Apparatus for Measuring Contact Potential Difference in Air," *Rev. Sci. Inst.*, **16**, 343–345, December 1945; J. A. Williams, "Crystal Driven Modulator for D-c Amplifiers," *Electronics*, **18**, 128–129, December 1945. See also references, Sec. 3·10, p. 44.

at 2615 cps, and there are four separate pickup probes or plates. The alternating current induced by the diaphragm vibration is approximately 1 per cent of the d-c potential, and the accuracy is better than 0.1 per cent.

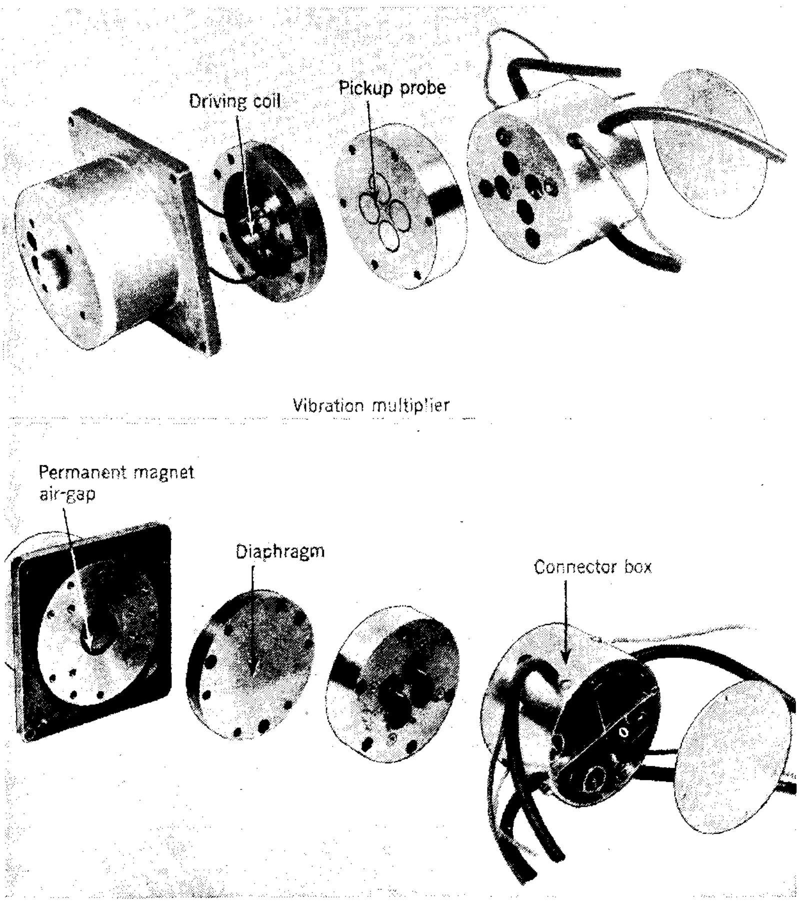


FIG. 15·24.—Vibrator multiplier, disassembled view.

An equating circuit that is capable of very high precision, particularly for laboratory use, is shown schematically in Fig. 15·26. A sensitive galvanometer, a lamp, a double photocell, and a difference-taking d-c amplifier are connected in such a way that a small current in the coil causes a mirror to move, changing the photocell balance and giving a large d-c output voltage. This is sometimes called the "Roberts galvanometer" and is discussed more fully in Vol. 20. This principle has been used in a circuit to regulate the beam current of a cyclotron.[1]

In applications where the peak value of an a-c signal is compared with a d-c reference, a diode demodulator (detector) may be used as a comparing circuit. The loading on the peaks of the output by the diode

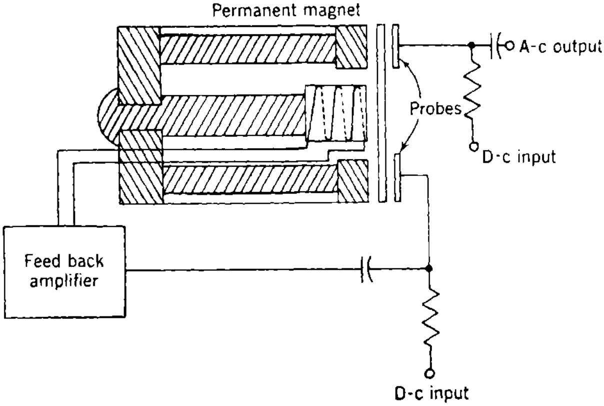


FIG. 15·25.—Modulation circuit employing vibration multiplier.

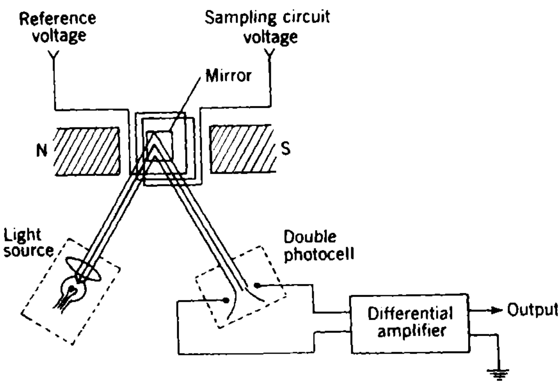


FIG. 15·26.—Use of galvanometer and photocells in comparison circuit.

detector may generate harmonics in the a-c output; such loading is minimized by using high-impedance circuits and by providing a low-impedance a-c source for the detector (see Figs. 15.27 and 16.20).

Summary of Comparison Circuits.—Some of the features of comparison circuits described in this section are summarized in Table 15·3.

[1] Ridenour and Lampson, "Thermionic Control of Ionization Gauge," *Rev. Sci. Inst.*, **8**, No. 5, 162–164, May 1937; G. Asset, "Photoelectric Galvanometer Amplifier," *Electronics*, **18**, 126–129, February 1945.

TABLE 15·3.—SUMMARY OF COMPARISON CIRCUITS

	Carrier frequency	Accuracy	Shock and vibration	Position effects	Temperature effects	Loading of reference voltage	Actuating voltage
Single-ended tube, low I_p*	.....	±150 mv	10 mv	None	None	Yes	..
Single-ended tube, high I_p*	.....	±300 mv	10 mv	None	None	Yes	..
Dual tube in symmetrical circuit, low I_p*	.....	± 50 mv	15 mv	None	None	No	..
Same high I_p (6SN7)*	.....	±100 mv	15 mv	None	None	No	..
Diode and triode:* Symmetrical modulators low current at balance†	Any	± 50 mv	15 mv	None	None	Depends on circuit	..
Brown converter	50–60 cps	± 1 μv at 10 ohms	Somewhat affected	Requires rebalance for best accuracy	None	Not at balance	6–9
Leeds and Northrup vibrator	60 cps	Less than 15 μv	Somewhat affected	Requires rebalance for best accuracy	None	Not at balance	35
Roberts galvanometer	.....	0.01 μa	Very serious	Serious	None	Small	..
Gen. Mot. Res. Lab. switch modulator	75 cps	10^{-9} v at 5 ohms	Not serious	Not serious	None	Not at balance	....

* For 10 per cent change in heater voltage, 100 hr. time.
† This refers to symmetrical circuits in which two cathodes are used.

15·5. Control Elements.—In a degenerative regulator the element that modifies the input in order to obtain the desired output may be called the control element. Modification of the input to obtain the desired output may be realized by attenuation introduced by a variable impedance or by a component added or subtracted from the input by a device such as a variable autotransformer.

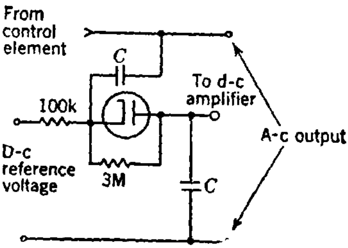


FIG. 15·27.—Diode detector equating circuit.

Variable impedance control elements may be resistive or reactive. Vacuum tubes, thyratrons, potentiometers, rheostats, and carbon pile elements are resistive control elements. Variable inductors in which impedance variation is accomplished by mechanically changing air gaps or by magnetic saturation are reactive control elements. Variable transformers provide means for adding to or subtracting from the input.

Vacuum tubes are used in series or shunt to regulate d-c supplies. Vacuum tubes used as control elements usually have fair linearity and very short time constants. Their high input impedance for the controlling voltage results in very high power gain. Their worst feature is low efficiency, sometimes less than 50 per cent.

The choice of a control tube for a specific application is influenced by a number of factors. The allowable plate power loss, maximum plate current, and minimum voltage for rated plate current are very important limitations. Less important are heater power loss, allowable heater-cathode potential, and the desirability of keeping the number of separate tube types as low as practical in a particular instrument.

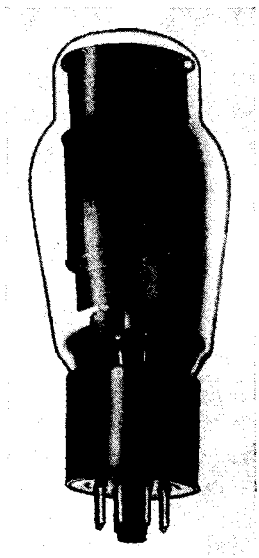


FIG. 15·28.—6AS7-G control tubes.

In general the series control tube should have high transconductance and high plate dissipation rating and should pass rated plate current at low plate-cathode potentials. Popular control tubes for direct current regulators include the 6V6, 6L6, 6B4, 28D7, 26A7, 6Y6, and 6AS7-G. The 6AS7-G (Fig. 15·28) is an excellent new tube of high current-rating (200 ma) and low voltage drop developed especially for series control use. Its low μ, 2.1, requires a rather large control voltage range from the amplifier, but this is not serious.

Circuits incorporating shunt control tubes are usually less efficient in regulator circuits than series control tubes, particularly for high-current regulators. A shunt regulator using a specially developed tube to regulate a high-voltage, low-current supply is shown in Fig. 16·54, Sec. 16·8, and is described in detail in Vol. 22.

Thyratrons can be used as controlled rectifiers to provide a regulated supply at higher efficiency than that obtainable with high vacuum tubes. Thyratron controlled rectifiers usually require extra filtering and shielding for satisfactory operation.

Saturable reactors[1] are the basis for simple a-c regulators widely used to provide reasonably constant voltage from power lines. Commercial versions are described in Sec. 16·3. The power efficiency of saturable reactor a-c regulators may be high compared with tube d-c regulators, as the variable impedance introduces reactive voltage drop instead of resistive (hence dissipative) drop. Saturable reactors are characterized by time constants ranging from milliseconds to seconds, depending on the

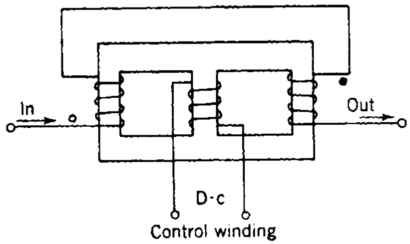


FIG. 15·29.—Saturable reactor with d-c control.

design and the driving impedance. In a typical case, a simple a-c regulator provides full correction after 2 cycles at 60 cps.

Saturable reactors, being very nonlinear, introduce harmonics into the output potential that may be undesirable and complicate the design of a regulator for a specific application.

Figure 15·29 is a sketch of a saturable reactor in which the introduction of the alternating current into the d-c control winding is minimized by a balanced construction. In most designs the d-c control winding has a large number of turns in order to obtain the maximum NI with a minimum current. The time constant resulting from this high-impedance control winding may be a disadvantage in some applications. It is minimized by pentode drive.

One simple application of a saturable reactor is in the so-called[2] "swinging choke" for power supplies.

A transformer with an adjustable slider electromechanically activated by a solenoid or a servo motor can be used as a control element. Motor-operated regulators in which transformer taps are changed are widely

[1] See Sec. 12·23.

[2] Regulation of rectifier power supplies can be improved by the use of an input choke in which impedance decreases with increase in load current. This tends to make the filter look more like a condenser input filter for higher currents, resulting in an improvement in regulation for load changes.

used in power applications where rough regulation with high efficiency is desired.

A carbon pile variable resistance element, in which resistance is an inverse function of pressure, is used as a control element for directly regulating high-current low-voltage d-c supplies (see Fig. 15·30). The carbon pile is actuated by a solenoid which samples the output, a spring acting as a reference element. Vibration of the mechanism of a carbon pile regulator and its relatively slow response limit the smoothness of output under practical operation conditions. A variable resistance, mechanically actuated by a servo motor, a solenoid, or other electro-

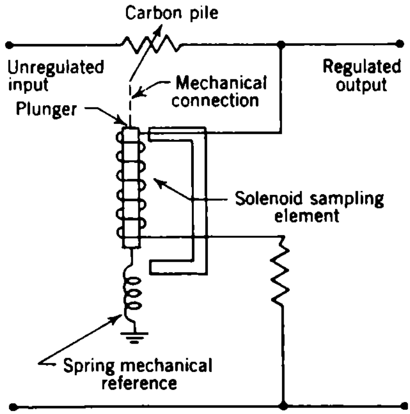


FIG. 15·30.—Carbon pile voltage regulator.

mechanical devices can be used as a regulator control circuit. Variable resistance elements have the advantage of controlling alternating current without wave shape distortion but are not so efficient as elements like saturable reactors. The primary disadvantage of variable resistance elements is the slow response due to the mechanical motion required to actuate the element. These characteristics are particularly troublesome in high-current units.

In many applications the use of an electromechanically actuated control element to provide slow control, combined with a vacuum tube or saturable reactor control element to provide a small amount of rapid control, may be of advantage. This dual arrangement may be more efficient and provide much better regulation than any single control element regulator.

CHAPTER 16

PRACTICAL REGULATOR DESIGN

BY A. JACOBSEN AND J. V. HOLDAM, JR.

16·1. Design Considerations.—Chapter 15 contains detailed information on the characteristics of the four major parts of regulators: reference elements, sampling circuits, comparing circuits, and control elements. This chapter is a study of the over-all characteristics of regulators, and an attempt is made to show how these characteristics influence regulator design.

In general, regulators are required for circuits or systems whose satisfactory operation depends upon constancy of voltage or current and that operate under conditions where there are variations in supply voltage and load. Since regulation is required for specific purposes, specifications can be written precisely. The type of regulator called for is determined by the input voltage, whether it is alternating or direct current, how much the input voltage fluctuates, and the required range and accuracy of output. The actual circuits used in the regulator are determined by more detailed specifications such as the type of operation: whether it is industrial or military; shipborne, airborne, or ground; mobile or stationary; or laboratory. Even more detailed specifications include such information as the magnitude of the output voltage and the per cent regulation required; the upper limit of the output impedance, usually determined by the range of load (current) variation; the degree of dynamic stability, determined by the character of the transients expected and the sensitivity of the circuit following the regulator to such transients; and the required static stability, usually determined by the length of time that the equipment must remain stable between adjustments. For the sake of classification, regulator specifications have been divided into two main groups: a-c regulation and d-c regulation. Each group is subdivided into several classes of regulators, the distinction between classes depending primarily upon the power level available from the regulator and whether the regulating action is simple or degenerative. Admittedly, this division is arbitrary; in some cases the classes overlap, and there are many omissions. It is impossible in the space available to include circuit diagrams, operating characteristics, and a theoretical discussion of all common regulator designs. Examples of most of the various classes of regulators are described, and many circuit diagrams and operating characteristics are included.

A-c Regulation.—Alternating-current regulation may be divided into four broad classes: the regulation of power generated by a large central station with well-stabilized frequency, regulation of a local a-c generator or inverter, simple regulation at low power levels for specific circuit applications, and regulation of the output of electronic or mechanical oscillators. Table 16·1 gives a résumé of the principal characteristics of these classes of regulators.

Simple regulators, employing magnetic saturation, are extremely useful in regulating a-c power, provided the frequency is constant. Commercial regulators, employing magnetic saturation, are available for a large range of power levels. Because of its importance as a class, this type of regulator is discussed in some detail in Sec. 16·3. Electromechanical regulators, to which Sec. 16·4 is devoted, are considerably more flexible. Motor control of a variable transformer provides a-c regulation for conditions that magnetic saturation regulators cannot meet. In particular, this method of regulation is relatively insensitive to fluctuations in frequency and output power factors, two conditions that must meet stringent requirements for satisfactory operation of magnetic saturation regulators.

Nonsaturating resonant circuits to provide constant current have sometimes been used, particularly for controlling the input to high-power oscillators. Application of this type of a-c regulation is limited to central station power sources with carefully controlled frequency, since the resonant circuit has a high Q and hence is sensitive to frequency variations. The control of the output of local inverters by means of varying the field current with a carbon pile variable resistor in a degenerative circuit has found considerable application in airborne and other portable inverters. Field control by carbon pile regulators is not too satisfactory but has found widespread use owing to the lack of a better system. These classes of regulators, together with simple regulators and oscillators employing nonlinear impedances as stabilizing elements, are discussed in Sec 16·5.

D-c Regulators.—Table 16·2 divides d-c regulators into several classes, the classes being determined primarily by the voltage, current, and power levels, and gives the general characteristics of each class. Field control of d-c generators is the usual practical way to provide d-c regulation at high power levels. Carbon pile control elements in a degenerative regulator circuit are widely used in controlling the field current in low-voltage, high-current generators. For the regulation of low voltage at medium current and power levels, the carbon pile can be used as a series impedance control element in a degenerative regulator circuit. This

TABLE 16·1.—A-C REGULATORS

Class	Power level, volt-amps	Method	$\frac{dE_{out}}{dE_{in}}$, %	Output impedance	Static stability	Dynamic stability	Efficiency
Central station power	15–5000	Magnetic saturation simple regulator	5	Low	Good	$\approx \frac{2}{f}$ sec	90%
Central station power	Up to 50,000	Motor control of variable transformer degenerative regulator	< 1	Low	Excellent	$\approx \frac{2}{f}$ sec	50–90%
Central station power	Up to 50,000	Resonant nonsaturating constant-current simple regulator	$\frac{dI_{out}}{dE_{in}} \approx 5$	$\frac{dI_{out}}{dZ_{out}} \approx 5\%$	Good	$\approx \frac{1}{f}$ sec	90%
Local inverter	500–2000	Field control, carbon pile, degenerative regulator	10	Low	Poor	$\approx \frac{1}{10}$ sec	75%
Simple regulator	Up to 10	Nonlinear impedance	10	High	Fair to poor	Good to poor	Low
Oscillators	Up to 5	Nonlinear impedance in degenerative regulator	< 1	High	Good	$\frac{1}{10}$ to $\frac{1}{2}$ sec	Low

TABLE 16·2.—D-C REGULATORS

Class	Power level, watts	Method	$\frac{dE_{\text{out}}}{dE_{\text{in}}}$, %	Output impedance	Static stability, sec	Dynamic stability, sec.	Efficiency
Low-voltage, high-current.	500–10,000	Field control carbon pile degenerative regulator	5	Low	Fair	$\sim \frac{1}{10}$	50–90%
Low-voltage, medium-current.	20–500	Carbon pile series impedance, degenerative regulator	5	Medium	Fair	$\sim \frac{1}{10}$–$\frac{1}{50}$	50–75%
Low-voltage, low-current.	1–20	Nonlinear impedance simple regulator	5	High	Fair	$\frac{1}{10}$–$\frac{1}{2}$	30–60%
Medium-voltage, medium-current......	0–5	Bridge regulator	1–5	High	Fair	$< \frac{1}{1000}$	30–50%
	0–10	Nonlinear impedance simple regulator	1–5	Medium	Good to poor	1–$\frac{1}{1000}$	Low
	0–50	Electronic degenerative regulator	0.01–1	Low	Good to excellent	$< \frac{1}{1000}$	Low
High-voltage, low-current............	0–5	Electronic degenerative regulator	1	High	Good	$< \frac{1}{1000}$	Low
	0–5	Stabilized oscillator, rectifier	1	High	Good	1–$\frac{1}{1000}$	Low

type of regulation is discussed in Sec. 16·4. Low voltage at low current can often be regulated to the required accuracy by the use of nonlinear impedances in simple regulator circuits. Nonlinear impedances are available in a wide range of operating characteristics. They can also be used in simple regulators for medium voltage at low current. These classes of regulators together with the regulating bridge circuits are discussed in Secs. 16·2 and 16·8.

To provide very precise regulation in the voltage range of 100 to 500 volts, electronic degenerative regulators are employed. Several useful circuits are described in Secs. 16·6 and 16·7. Section 16·2 contains some theoretical considerations of the characteristics of degenerative regulators that are very useful in designing such circuits. Section 16·8 includes some regulators designed to operate at high voltage and low current. Volume 22 of this series includes considerably more information on high-voltage regulated supplies.

Combination of Regulators.—Although the various regulators discussed in this chapter are complete units, it should not be inferred that improved performance cannot be achieved by a combination of separate regulators. Indeed, for many applications, using two or more regulator circuits in cascade gives greatly increased performance. Regulator circuits, however, may not be put in cascade indiscriminately. In general, the use of two or more regulators in series improves the output of simple regulators; this is so because simple regulators contain only passive elements and tend to divide the input fluctuations by a definite fraction. On the other hand, degenerative regulators are characterized by active elements and are limited by the drift and change in characteristics of the active elements.

The use of simple regulators in series can improve only the regulation; the output impedance is either unaffected or made worse as is the dynamic and static stability. For example, two a-c voltage regulators of the magnetic saturation class will give a fractional regulation that is approximately equal to the product of the regulation fractions of the individual regulators. The output impedance, however, is the same as for one regulator; the static stability is somewhat poorer, whereas the dynamic stability is somewhat improved; and the efficiency is the product of the efficiencies of the two individual units.

If regulator circuits are combined in other ways than by cascading, improved performance can be gained in degenerative regulators as well. For instance, Sec. 16·6 includes the description of a degenerative regulator that employs two comparing circuits and two reference elements. One of the circuits, wholly electronic, is designed to give the regulator a very short response time and hence good dynamic stability. The electronic reference element, however, does not have adequate static stability, and this is provided by a separate comparing circuit utilizing a mechanical converter. The second, more accurate, comparing circuit is used to correct the output of the first and, hence, gives the regulator excellent static stability.

No general rules can be laid down to follow in combining regulator circuits. Combinations will be useful wherever input fluctuations are very large and where the specifications on regulation and static and dynamic stability are very severe.

16·2. Prediction of Performance.—The basic purpose of the present chapter is to enable the designer to choose the appropriate elements to meet the desired specifications. This section contains theoretical treatments of idealized circuits so the designer will have a better understanding of the workings and limitations of the various circuits.

Since the typical circuits that are analyzed are idealized, it is rarely sufficient to predict performance wholly on the basis of calculations; it is usually necessary to use laboratory methods to make the final adjustment of variable parameters and to check the performance.

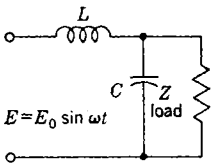


FIG. 16·1.—Resonant constant-current circuit.

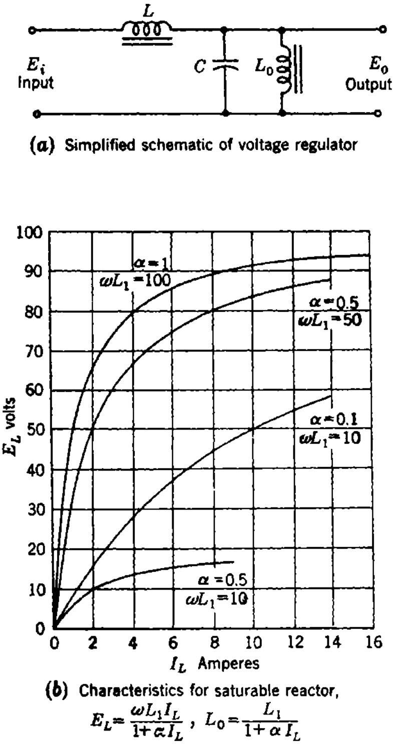


FIG. 16·2.—Simplified circuit and characteristics of reactor for magnetic saturable reactor a-c voltage regulator.

Simple A-c Regulators.—There are a great many applications, both industrial and laboratory, where regulated a-c voltage or current is required. One of the simplest regulators, the resonant *constant-current* regulator, is shown in Figure 16·1. It is easy to show that the average load current, $\bar{I}_Z$, is equal to $2E_o/\pi\omega L$, where $E_o \sin \omega t$ is the input from a zero-impedance source and $\omega = 1/LC$. *Thus the load current is independent of the load resistance under these conditions.*

Since they have found wide application, *constant-voltage* units are manufactured commercially by many manufacturers (see Sec. 16·3). Although commercial designs differ, they can all be understood from the explanation of a simplified circuit.

In Fig. 16·2 E_o is independent of E_i provided L_0 has a characteristic of the general form $L_0 = L_1/(1 + \alpha I)$ and L and α are properly chosen.

If all the components are without resistance, the voltage relations can be written very compactly using only elementary circuit theory.

$$E_i = E_o + \frac{L}{L_o}(1 - \omega^2 L_o C)E_o; \tag{1}$$

L_o is a saturable reactor whose performance can be characterized by the empirical relation

$$L_o = \frac{L_1}{1 + \alpha I}, \tag{2}$$

where L_o is the equivalent inductance, and I is the magnitude of the current through L_o. Thus, L_o is a real number. Also,

$$|I| = \frac{E_o}{\omega L_o}. \tag{3}$$

I and L_o can be eliminated from Eqs. (1), (2), and (3), giving

$$E_i L_1 = -\frac{\alpha}{\omega} E_o^2 (1 - \omega^2 LC) + E_o \left[L(1 - \omega^2 L_1 C) + L_1 + E_i \frac{\alpha}{\omega} \right]. \tag{4}$$

But $LC = 1/\omega^2$, so (5)

$$E_o = \frac{E_i \omega L_1}{L\omega + \alpha E_i} \tag{6}$$

If $L\omega \ll \alpha E_i$, Eq. (6) becomes (7)

$$E_o = \frac{\omega L_1}{\alpha} \tag{8}$$

and E_o is independent of E_i.

This is only approximate, for the assumptions regarding the behavior of L_o are somewhat drastic. It is also noted that the above derivation has neglected the load current. The approximation improves as E_i increases. As will be seen later (Sec. 16·3), balancing terms may be introduced to give more nearly perfect regulation by suitable simple connections.

This regulator has essentially the same stability characteristics as the constant-current regulator. Dynamic response is approximately $2/f$ sec, and static stability is limited by the stability of the circuit elements.

Simple D-c Regulators.—Circuits for regulating direct current with wholly passive elements (simple regulator) usually consists of linear and nonlinear impedances combined to provide reasonably constant output. Detailed characteristics of the most popular nonlinear impedance are given in Sec. 15·2 (Reference Elements).

Most nonlinear impedances, of which a glow tube is a good example, are connected as shown in Fig. 16·3 to form simple regulators. The regulation and internal impedance are calculable from the values of the linear elements and the data on nonlinear elements contained in Sec. 15·2. For example, in the circuit shown in Fig. 16·3, if the glow tube is a VR-105, the average dynamic impedance is 40 ohms. Table 15·1 lists a VR-105 as having a 0.7 per cent increase in E_g over the useful current range of 10 to 30 ma. Hence

$$R_G = \frac{\Delta E_G}{\Delta I_G} = \frac{0.007 \times 105}{0.030 - 0.010} = \frac{0.8}{0.02} = 40 \qquad \text{ohms}, \tag{9}$$

and

$$\text{per cent regulation} = \frac{100\ \Delta E_L}{\Delta E_o} = \frac{100 \times \dfrac{40R_L}{40 + R_L}}{R_1 + \dfrac{40R_L}{40 + R_L}} \cong \frac{40}{R_1 + 40}\,100 \tag{10}$$

if R_L is large compared with 40 ohms.

Such calculations are only approximate because the value of R_G is an average value. If dE_G/dI_G is measured as a function of I_G, it is not

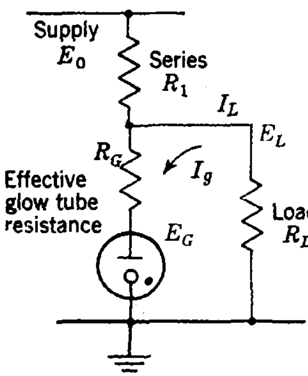


FIG. 16·3.—Simple glow tube regulator.

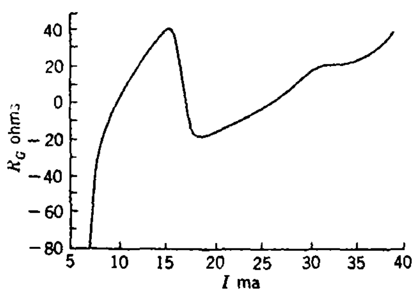


FIG. 16·4.—Internal impedance of VR-105 vs. current.

constant; in fact it fluctuates rather violently. A typical example is shown in Fig. 16·4.

For constant average current $\bar{I}_G$ the impedance is not independent of the frequency of fluctuations of I_G. The reactive component depends on $\bar{I}_G$ in an erratic fashion that does not lend itself to critical analysis. Figure 16·5 shows typical R_G vs. frequency curves for VR-105's.

No analysis is made of circuits employing other nonlinear impedances. Because of their temperature instability other nonlinear impedances have not found so wide applications as glow tubes.

Perhaps the most widely used simple d-c regulator is a condenser. It is a common practice for designers to place condensers across voltage buses whose stability is impaired by load transients. Capacity across a

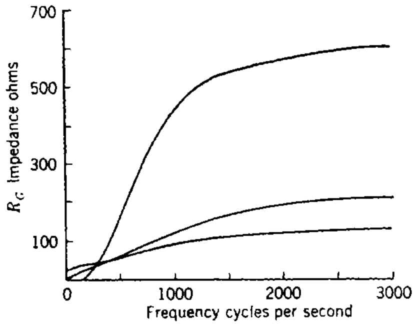


FIG. 16·5.—Internal impedance of VR-105's vs. frequency.

source is particularly effective if the source impedance is large. Figure 16·6 is a schematic of such an arrangement, where R is the source impedance, R_L is the load impedance, and C is the filter, or regulating, capacitance. Such a regulator has an internal impedance of $R/\sqrt{\omega^2R^2C^2 + 1}$. The d-c regulation dE/dE_s is obviously $R_L/(R + R_L)$; whereas the regulation for fluctuations is

$$\frac{dE}{dE_t} = \frac{\sqrt{K^2 + \omega^2}}{RC(K^2 + \omega^2)}, \tag{11}$$

where

$$K = \frac{R + R_L}{R_L}\frac{1}{RC}. \tag{12}$$

Electronic Degenerative Regulators.—A simple form of degenerative regulator is the d-c cathode follower (Fig. 16·7). From the definition

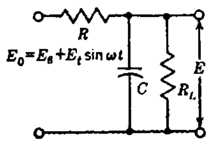


FIG. 16·6.—Regulation with capacitor shunting load.

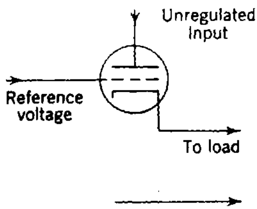


FIG. 16·7.—Cathode-follower degenerative regulator.

of μ and g_m it is apparent that the per cent regulation is $1/\mu \times 100$ and the internal impedance is $1/g_m$. If a pentode is used, the regulation is very good, since the plate current is essentially independent of the plate voltage. Pentodes are often difficult to use, however, since they require stabilized screen voltages in order to obtain the pentode benefits.

The transconductance bridge (Fig. 16·8) has excellent regulation but very high internal impedance. Balance is achieved when

$$g_m = \frac{R_1 + R_2}{R_2 R_3}. \tag{13}$$

The bridge may be unbalanced to give either negative or positive regulation.

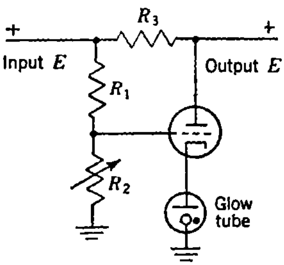


FIG. 16·8.—Transconductance bridge.

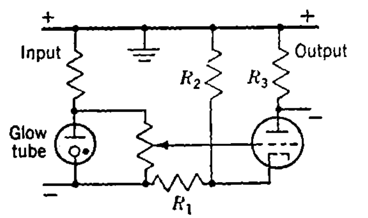


FIG. 16·9.—Degenerative bridge regulator.

The amplification bridge (Fig. 16·9) is similar to the transconductance bridge in that regulation is good but output impedance is high. The condition for balance is approximately

$$\mu = \frac{R_1 + R_2}{R_1}, \tag{14}$$

provided R_1 is small enough so that $R_1 i_p < R_1/(R_1 + R_2)E_{in}$.

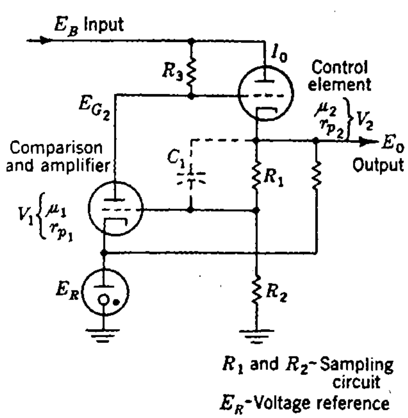


FIG. 16·10.—Degenerative d-c regulator.

The most popular form of electronic degenerative regulator is shown in Fig. 16·10. The following analysis neglects the effect of the internal impedance of the reference element.

Connecting R_3 to E_B has been found desirable, as the amplifier V_1 then operates in a more linear manner and the gain is somewhat higher than if R_3 were connected to E_o. Using current from the regulated output, as in Fig. 16·10, to keep the glow tube operating is best practice but requires more control tube capacity. To find regulation, let

$$N = \frac{R_2}{R_1 + R_2}, \qquad G_1 = \frac{\mu_1 R_3}{r_{p_1} + R_3}, \qquad K = \frac{r_{p_1}}{r_{p_1} + R_3}. \tag{15}$$

$$\text{Regulation} = \frac{dE_o}{dE_B} = \frac{\mu_2 K + 1}{1 + \mu_2(1 + G_1 N) + \frac{I_o r_{p_2}}{E_o}}. \tag{16}$$

If $G_1 N \gg 1$, $\mu_2 > 1$,

$$\frac{E_o}{I_o} > r_{p_2}, \tag{17}$$

then

$$\text{regulation} = \frac{dE_o}{dE_B} \cong \frac{\mu_2 K + 1}{\mu_2 G_1 N}. \tag{18}$$

The output impedance (sometimes called source impedance) is given by

$$r = \frac{r_{p_2}}{1 + \mu_2(1 + G_1 N)}. \tag{19}$$

If $G_1 N \gg 1$, $\mu > 1$, then

$$r \cong \frac{1}{g_{m_2} G_1 N}. \tag{20}$$

If the regulator is fed through a source impedance R_s (rectifier, filter, etc.) and has regulating factor R against changes in input voltage, then $r_t = RR_s + r$.

If the preceding approximations are considered,

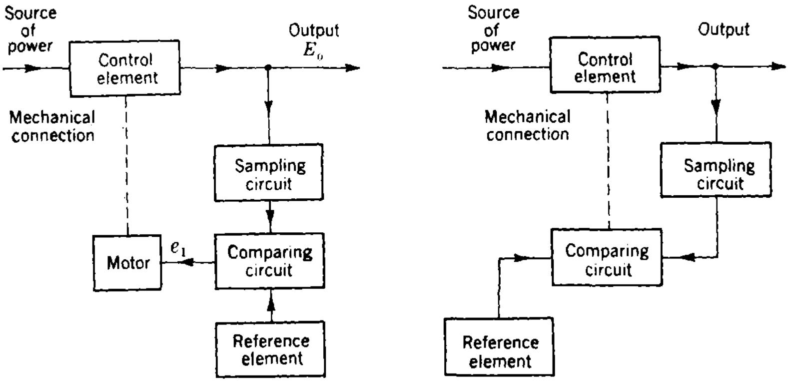
$$r_t = \frac{(\mu_2 K + 1)R_s + r_{p_2}}{\mu_2 G_1 N}. \tag{21}$$

In predicting the estimated performance of a regulator more precisely than by the preceding analysis, an experimental check on the behavior of the supply source E_B with I_o should be made for the line and load variations. From the tube characteristics for the control tube or tubes, the grid-cathode voltage excursion for the plate voltage and current can be determined.

The next step is to determine the gain necessary to provide the required control tube grid-cathode voltage variation for the allowable difference between regulator output and voltage reference. The comparing circuit and amplifier may then be designed with due regard to d-c levels and linear operation of the amplifier tubes.

Stability Considerations.—In addition to correct d-c conditions degenerative regulators require dynamic stability for satisfactory operation. Dynamic instability appears as oscillation of the feedback loop if the gain is more than unity when the phase shift is 180°. A lesser degree of instability may still cause excessive response to transients. The output impedance of the degenerative regulator also becomes high when the gain goes down and the phase shifts in a direction to make the loop less stable. In general, a wide bandwidth of the amplifier improves the dynamic stability, provided phase shift remains small over the pass band. Chapters 9 to 11 deal with the necessary stability theory.

The long-time stability of regulators is primarily determined by the effects of aging on the reference element and comparing circuit. This subject is discussed in Sec. 15·2 (Reference Elements) and Sec. 15·4 (Comparing Circuits).



(*a*) Motor actuated control element (*b*) Carbon pile control element

FIG. 16·11.—Mechanical degenerative regulators.

Mechanical Degenerative Regulators.—This class of degenerative regulator is considered separately because mechanically actuated control elements may differ in many important characteristics from electronic control elements. As indicated in Tables 16·1 and 16·2 motor-driven control elements for alternating current and carbon pile control elements for direct current are about the only satisfactory mechanisms for high-power regulation. Carbon pile control elements are used for direct variable attenuation of the power source voltage or for variable attenuation of field current in a generator.

Figure 16·11 shows block diagrams of the two principal types of mechanical degenerative regulators. The quality of the static regulation of the circuit of Fig. 16·11*a* is determined by the change in output voltage required to cause the motor to operate. As an example, let the output E_o be 100 volts, the voltage reference be a 1-volt standard cell, and a switch modulator be used for the comparison circuit.

If the amplifier gain is 10,000 and 30 volts output is required to break the starting friction of the servo motor, then with the loss in gain of 100 in the sampling circuit the ΔE_o necessary to make corrections is

$$\Delta E_o = \frac{\text{sampling circuit loss} \times \text{volts to operate motor}}{\text{amplifier gain}}$$

$$= \frac{100 \times 30}{10{,}000} = 0.30 \quad \text{volts.} \qquad (22)$$

This is ± 0.3 per cent, if the output is 100 volts. To realize a dead space of 0.1 per cent would thus require three times this gain. The gain of the amplifier could be increased, the loss in the sampling circuit could be reduced (by using a higher reference voltage), and/or a more sensitive motor circuit could be used to achieve a higher static accuracy. The static stability limitations of reference voltage and equating circuit

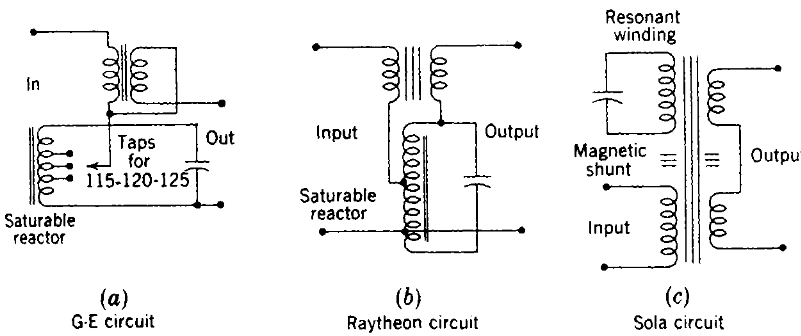


FIG. 16·12.—Commercial constant-voltage transformers.

are the same as for purely electronic regulators. Dynamically these regulators, like all electromechanical servos, have a time lag dependent on the motor and mechanism and may be dynamically stabilized by the usual phase-advance networks or tachometers. (*Cf.* Chaps. 9 to 11).

16·3. Regulation Using Magnetic Saturation Constant-voltage Transformers.—As explained in the preceding section voltage regulators can be built that depend on the saturation of a magnetic material to produce the necessary nonlinear impedance. Figure 16·12 shows schematics of three commercial versions of constant-voltage transformers.[1] Although the circuits are different in detail, they all operate on essentially the same principle. There are two respects in which these circuits differ from that Fig. 16·2 (Sec. 16·2): The output is taken from only part of L_0, and the output E_o is the difference between the voltage from L_0 and the voltage across L.

Since in Fig. 16·2, $E_o = \omega L_1/\alpha$, L_1 must be fairly small to keep E_c to the proper value, and consequently C must be large ($L_1 > L$). If the

[1] There are many possible combinations of reactors, transformers, and capacitors that achieve the desired simple regulator effects. For other see K. J. Way, "Voltage Regulators Using Magnetic Saturation," *Electronics*, July 1937; also F. Terman, *Radio Engineers' Handbook*, McGraw-Hill, New York, p. 615.

output is taken from a part of L_0 then $E_o = K(\omega L_1/\alpha)$, and L_0 can be made correspondingly larger.

As shown in Eqs. (6) to (8), the system does not give perfect regulation for practical values of α. However, a perfectly flat (or even a drooping) E_o vs. E_{in} characteristic can be obtained by connecting the output of L_0 in opposition to a small secondary winding on L. This also allows a larger L_0, since the output voltage is reduced still further. By giving the secondary of L the proper turns ratio the bucking voltage just cancels the E_{in} term of the voltage taken from L_0.

For many applications the use of these regulators provides the most economical and convenient way of realizing satisfactory operation of electronic instruments. However, the output waveform is necessarily distorted by core saturation, and this may be serious in the operation of some instruments.

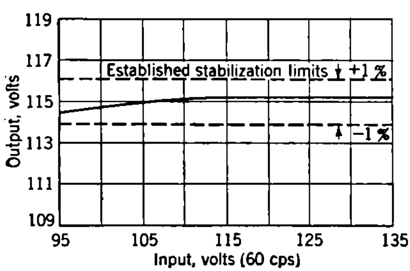


FIG. 16·13.—Output voltage vs. input voltage for 500-volt-ampere unit at rated unity power factor load. Data subject to 1 per cent manufacturers' tolerance.

Probably the most serious disadvantage of these simple regulators is the large error caused by frequency change. This is due to the use of resonance to achieve regulation. The percentage change in output potential is approximately proportional to the percentage change of frequency. This is satisfactory for large interconnected power systems where frequency is within $\pm \frac{1}{4}$ of a cycle, but the error may be very serious for small isolated power systems such as are used in ships and some industrial plants and small communities. For aircraft power systems the errors due to frequency change would be prohibitively large.

A test was run on two Sola 120-watt regulators in series to determine if cascade operation would be of value. For a line voltage change of 100 to 125 volts with a 40-watt load on the second regulator, the first regulator output changed approximately 1 volt while the second regulator output changed less than 0.1 volt. The 40-watt load was removed, and the output increased by 0.1 volt. The wave shape from the second regulator was somewhat more distorted than the output of the first, but not by more than a few per cent. This cascade arrangement would not be any better for changes in frequency than one regulator.

The following information relates to a General Electric Company regulator but is typical of the performance of all commercial units.[1]

The rated output voltage, when operating at rated load and rated

[1] Data taken from *GE Bull.* GEA-3634A.

(unity) power factor, varies less than ±1 per cent. Figure 16·13 shows the relation between input voltage and output voltage at rated load and unity power factor for a typical 500-volt-ampere stabilizer. The actual variation of the output voltage from the rated value is considerably less than ±1 per cent.

The variation in output voltage resulting from change in load with unity power factor, from no load to full load, with the input voltage constant, is less than ±1 per cent. For simultaneous variations of input voltage and load conditions, within the previously described limits, the output-voltage variations are less than ±2 per cent. Figure 16·14 shows the relation between input voltage and output voltage for various loads at unity power factor.

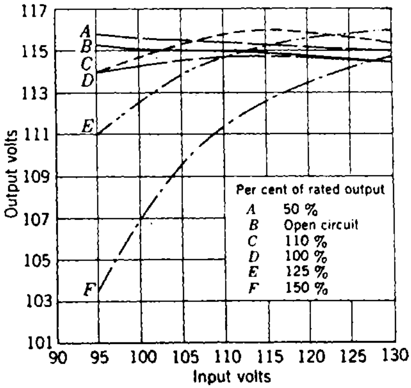


FIG. 16·14.—Typical regulation curves for various unity power factor loads for a standard 500-volt-ampere voltage stabilizer.

The output-voltage relations to load power factor and to load are shown in Figs. 16·15 and 16·16 respectively. Note that load power factors lower than unity decrease the output-voltage level. This

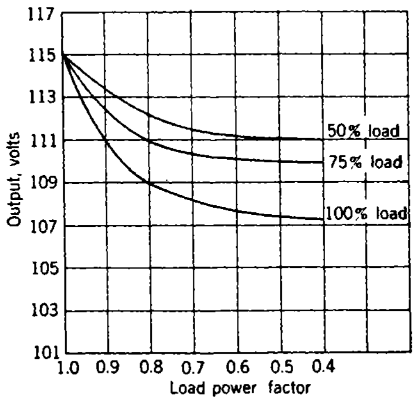


FIG. 16·15.—Output voltage vs. power factor at various loads for a standard voltage stabilizer.

decrease is considerably less for 50 per cent load than for 100 per cent load. The lowered output-voltage level, however, remains practically constant with change in line voltage for any given load or load power-factor condition.

A change of 1 per cent in the frequency of the input voltage causes a $1\frac{1}{2}$ per cent change in the output voltage, the voltage change being in the same direction as the frequency change. This is a very important characteristic and a serious limitation in many cases.

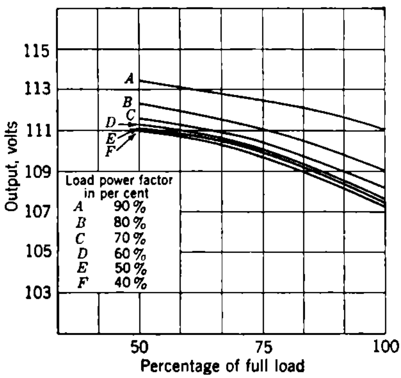


FIG. 16·16.—Output voltage vs. percentage of full load at various load power factors for a standard 500-volt-ampere voltage stabilizer. Input, 115 volts; nominal output, 115 volts.

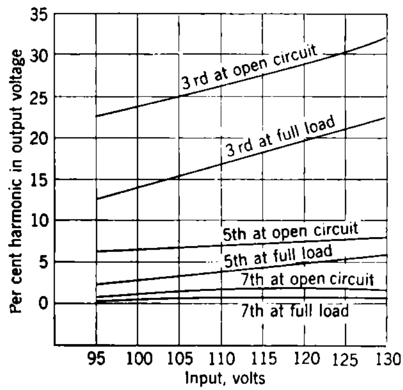


FIG. 16·17.—Harmonic content of standard voltage stabilizer.

Figure 16·17 shows the harmonic content of the output voltage for 115 volts input, full load. Special stabilizers incorporating a harmonic filter are available for applications that require low-harmonic content.

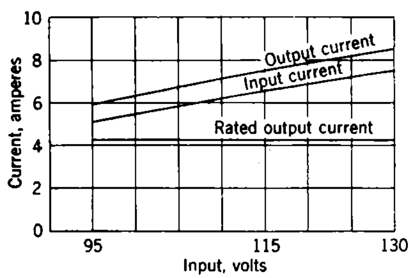


FIG. 16·18.—Input and output current under short-circuit conditions as a function of input voltage for a standard 500-volt-ampere voltage stabilizer.

An inherent characteristic which makes saturable reactor regulators very useful is the limited short-circuit current. At 130 volts input the current is limited to approximately 200 per cent of the rated output current. This characteristic protects both the unit itself and the load from excessive damage due to a short-circuit condition. This characteristic is very useful for applications that require large starting current such as motors. Figure 16·18 shows both input and output current for a 500-volt-ampere unit under rated and short-circuit conditions.

Table 16·3 is a summary of characteristics of a number of available units. Many other models are made for other voltages, frequencies, and incorporating output filters for special applications. The simplicity of construction of a typical model is indicated by Fig. 16·19, showing the interior of a Ratheon VR-4 250-volt-ampere unit.

Degenerative Voltage Regulation.—Simple a-c regulators, discussed in the first part of this section, are very frequency sensitive because the

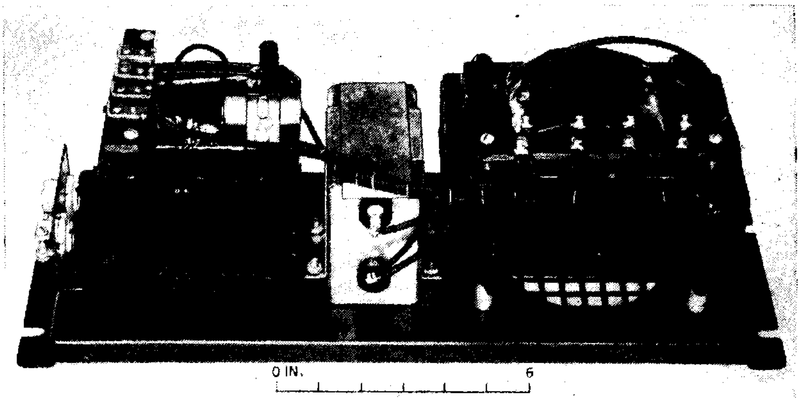


FIG. 16·19.—Interior of Raytheon VR-4 250-volt-ampere regulator.

regulating action is closely allied to a resonance phenomenon. However, magnetic saturation may be satisfactorily used to control an a-c voltage over a frequency range of one octave when employed in an electronic-degenerative circuit. Figure 16·20 shows such a circuit. This regulator was developed to regulate accurately a few watts of power for an airborne computer.

The output of this regulator is compared with a d-c signal derived from the 250-volt regulated bus. The output is maintained at 115 volts within 0.1 per cent for an input of 115 volts ± 10 per cent and for a frequency range of 380 to 450 cycles. The output wave shape is very good, with less than 0.1 per cent second harmonic and less than 0.2 per cent third harmonic when operated from an ordinary aircraft inverter. The output changes by less than 0.1 per cent for a 20 per cent change in output load impedance. Since the time constant for making corrections is of the order of $\frac{1}{8}$ sec., subharmonics of the input frequency are not removed either by the filter or by the regulator control action. This is a practical difficulty of some importance, since appreciable amounts of subharmonics are present in aircraft a-c systems.

In operation this degenerative a-c regulator depends on the use of a saturable reactor as the variable attenuator. The input transformer raises the input voltage; the saturable reactor provides variable attenuation determined by the direct current from the control amplifier; and the filter reduces the harmonics of the input wave and those introduced by the saturable reactor to very small percentages. A degenerative feedback loop is provided by a sampling diode detector, differential comparison amplifier, and control amplifier to keep the output very close to the desired value.

TABLE 16·3.—DATA ON TYPICAL STOCK VOLTAGE STABILIZERS, 60-CYCLE, SINGLE-PHASE, 95–130 VOLTS INPUT

Manufacturer	Output E, volts	Volt-amperes	Weight, lb
General Electric Co.	115	50	16
	115, 120, 125	100	18
	115, 120, 125	250	40
	115, 120, 125	500	68
	115, 120, 125	750	100
	115, 120, 125	1000	118
	115, 120, 125	2000	210
	115, 120, 125	3000	310
	115, 120, 125	5000	500
Raytheon Manufacturing Co.	6.3 or 7.5	30	8
	115	30	8
	115	60	18
	115	120	26
	115	250	46
	115	500	70
	115	1000	140
	115	2000	200
Sola Electric Co.	6.0	15	4
	115.0	15	4
	6.3	17	4½
	115.0	15	4½
	6.3	15	2
	115.0	15	2
	115.0	30	11
	115.0	60	12
	115.0	120	15
	6.0	25	10
	6.3	50	12
	115.0	250	28
	115.0	500	40
	115.0	1000	115
	115.0	2000	175

The loop is stabilized and filtered by C_1 and C_2. The use of a negative feedback loop around the amplifier provided by C_1 is of advantage in realizing good dynamic stability. It was found that C_2 aided in minimizing harmonics generated by the saturable reactor. The M-derived filter using chokes with large air gaps is an important feature in realizing very low harmonic content in the output waveform. The high-impedance diode detector is a factor in minimizing the harmcnics and in realizing stable operation.

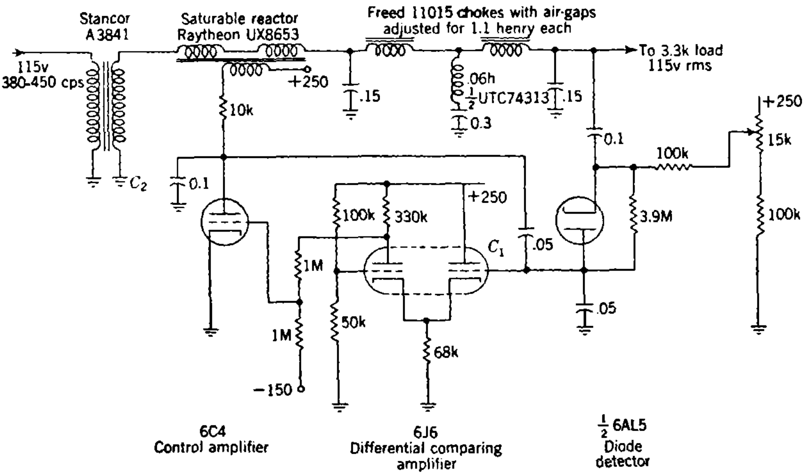


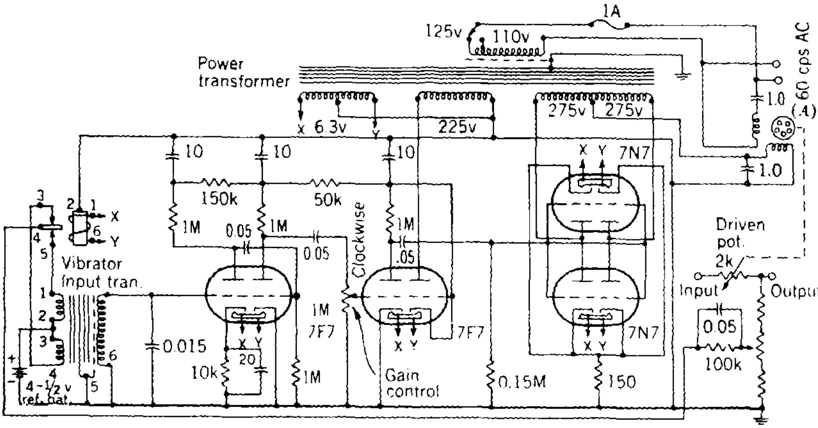
FIG. 16·20.—Degenerative a-c regulator using magnetic saturation.

The most serious effect of temperature on the operation of this circuit is in the increase of leakage current in the diode detection circuit. The leakage resistance of many good quality condensers of 0.1 to 0.01 μf may change by 25 megohms when the temperature is increased from 25° to 80°C. This is a very serious problem, as the diode detector circuit impedance must be made high, so a variation of 25 megohms will cause an error of about 0.4 per cent in this circuit, which has an impedance of about 100,000 ohms. Heater cathode leakage resistance in the diode is an important matter in some instances as well as leakage resistance of wiring, terminal boards, and sockets.

The order of stability required calls for the use of wire-wound resistors for the sampling and level setting circuits and a balanced comparison circuit.

16·4. Electromechanical Regulators.—An electromechanical degenerative regulator is a regulator in which the control element is mechanically actuated, for example, by a servomechanism. Mechanically actuated control elements may be variable resistances (rheostats, potentiometers, and carbon pile elements are good examples) or variable transformers in which a slider picks off the desired voltage or in which the coupling between input and output is varied. The degenerative electromechanical regulator requires the usual reference element, sampling, and comparing circuits in addition to the electromechanical servomechanism that provides the amplification and mechanical drive for the control element.

Alternating-current regulators using variable transformers for control elements have high efficiency and very large power-handling capacity.



(A) Minneapolis-Honeywell motor and gear box M623AY3X1 60 cycle 162 rpm with 2k pot. 2ϕ induction motor.

FIG. 16·21.—Brown Instrument Co. servo-operated regulator.

In d-c regulators the efficiency is low, as control is accomplished by variable series dissipation; however, the power capacity may be large. Both high efficiency and large power capacity are realized in the case of a generator with the output regulated by field control.

Electromechanical regulators employing motor-operated control elements usually exhibit discontinuities in their output-input characteristics. Such discontinuities arise from the "dead space" of the motor-control system. The dead space, in terms of volts, is the potential difference at the comparing circuit required to start the motor. In practical circuits the dead space cannot be eliminated, since reduction to zero requires infinite gain. As a result of the existence of the dead space the output characteristic is best defined by the change in output voltage required to make the motor operate.

Servo-operated Regulator.—Figure 16·21 shows a servo-operated control-element degenerative regulator. This example is a precision voltage follow-up position servo modified to function as a voltage regulator.

The unit is made up of a Brown Instrument amplifier and Minneapolis-Honeywell motor and gear box. The gear box has been modified by the addition of a dial.

The basic parts characteristic of degenerative regulators are easily recognizable: a series resistance control element, a resistance sampling circuit, a Brown Converter vibrator comparing circuit, a dry battery reference element, and an a-c amplifier. The two parallel-connected 7N7 dual-triode tubes provide a power phase-detector whose output is applied to one phase of a two-phase induction motor which drives the control element through a gear box. The 50- to 60-cycle carrier is applied to the vibrator, the phase detector, and one phase of the motor to synchronize the operation of the circuits.

The vibrator compares the reference and sampling circuit voltages and produces alternating voltage whose phase and amplitude is a measure of error. The input transformer supplies gain by efficient impedance matching. The three-stage a-c amplifier provides high gain, and the gain control allows adjustment for optimum dynamic performance. The sampling circuit contains only low-temperature-coefficient precision wire-wound resistors.

The unit is built to operate on 50 to 60 cycles, 110 volts. Similar equipment is available for operation on 25 to 40 and 350 to 480 cycles 110 or 220 volts and 22 to 28 volts direct current. The control of a d-c source is accomplished by a series resistance element, a 2-k potentiometer, and requires a minimum current through the potentiometer to obtain the necessary voltage drop. The maximum allowable current is limited by the potentiometer construction to approximately 30 ma. The dead space for this circuit is of the order of ± 50 μv, and with a $4\frac{1}{2}$-volt battery the accuracy obtainable would be of the order of 0.001 per cent. However, pickup in the sampling circuit and dynamic instability of the servo loop limit the useful amplifier gain and reduce the accuracy to about 0.01 per cent. As the battery reference may be in error by 0.05 per cent due to small temperature changes and aging, the static stability is about 0.05 per cent. The resistance sampling circuit requires very high quality components to make the sampling errors smaller than the reference element errors.

The rate at which correction can be made is primarily determined by the motor and gear mechanism. At full speed the potentiometer turns at 162 rpm requiring about $\frac{1}{2}$ sec to obtain correction equivalent to full variation of the potentiometer. This regulator may be used to provide long-time stability to a smoothed and partially regulated source, the voltage drop across the potentiometer being made just adequate to provide the required correction range so that the roughness due to wire voltage difference is small compared with the desired voltage accuracy.

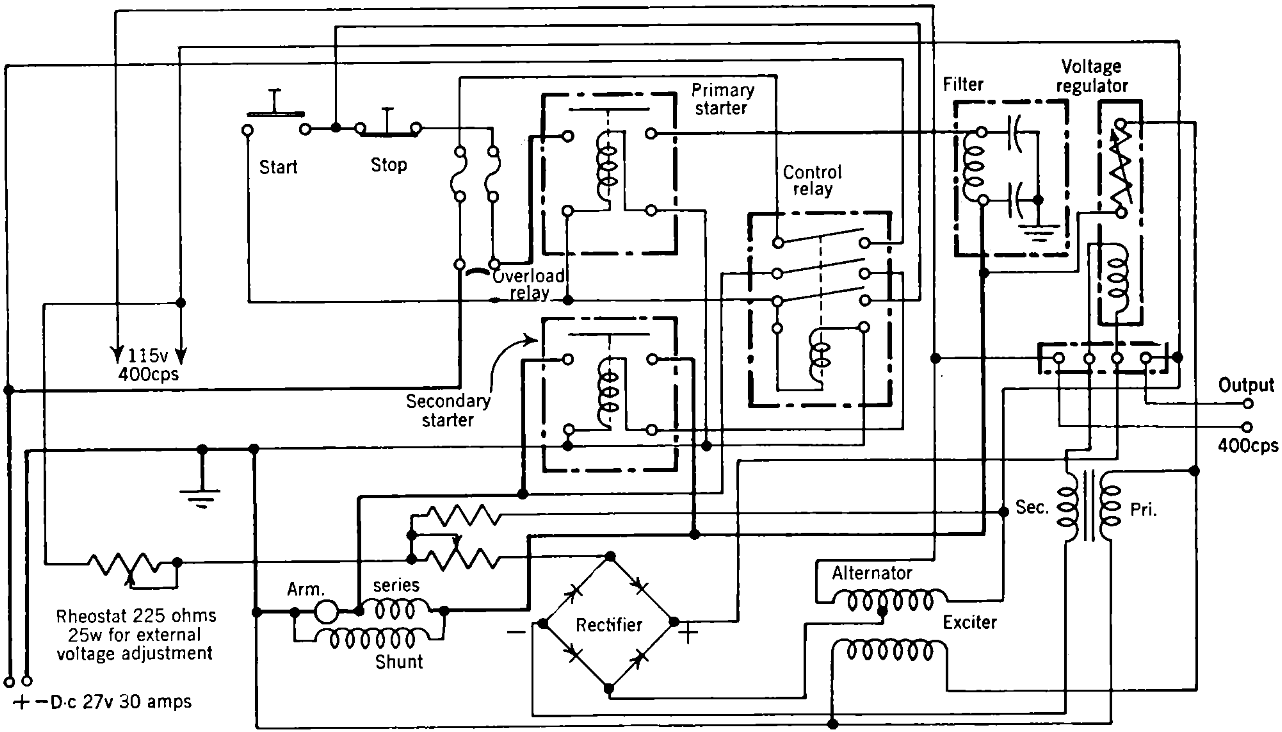


FIG. 16·22.—Wiring diagram for Leland aircraft inverter.

Many servomechanisms of Part II can be used in conjunction with suitable sampling, comparison, and control circuits to make an electromechanical regulator. This particular example can be used with an a-c sampling detector and possibly other control elements to regulate alternating current. The substitution of a standard cell as a reference element in place of the dry battery improves the static stability and makes better use of the accuracy of the mechanical vibrator comparing circuit.

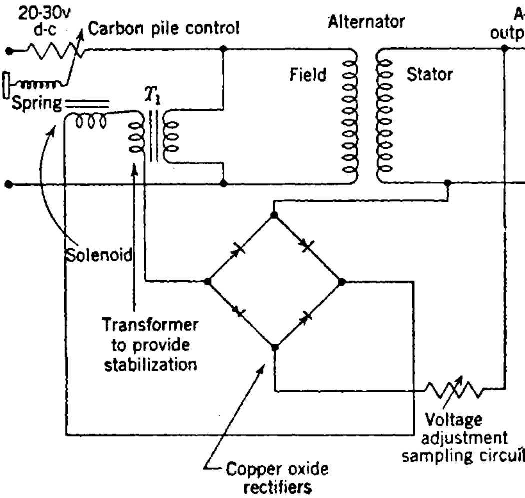


FIG. 16·23.—Simplified circuit of alternator with carbon pile regulator.

Carbon Pile Inverter Regulator.—In many applications of electronic instruments the prime power source is direct current, and a rotary inverter provides an economical means of obtaining a-c power. Field control provides a satisfactory way of regulating the output alternating current. Since the field current is large for reasonable size inverters, electronic control is impractical and carbon pile control elements have found general acceptance. The example given here is the field control of an aircraft inverter by a carbon pile degenerative regulator.[1] However, carbon pile regulators are used extensively to control the field current of d-c generators and as variable series resistors in low-power, low-voltage regulators. By proper arrangement of the sampling circuit either current or voltage can be regulated.

The performance specifications for the inverter shown in Figs. 16·22 and 16·23 are: input, 25 to 28 volts direct current; output, 115 volts alternating current, ±2 per cent, 0 to 13 amp; nominal frequency, 400 cps. The unit shown in Fig. 16·22 is made by Leland Electric Company.

[1] Inverters are discussed in detail in Vol. 17, Sec. 12·8.

Similar units are made by General Electric Company (P.E. 218D), Russell Electric Company (P.E. 218E), and others.

The wiring diagram is complicated by several relays and terminal boards but may be described simply in the following manner (refer to Fig. 16·23): Mounted on the same shaft are a compound wound d-c motor (not shown) and an alternator. The field for the alternator is excited from the d-c bus through a carbon pile resistance control element. The output of the alternator is sampled by a resistance voltage adjustment and a copper oxide bridge rectifier providing current to operate the solenoid of the carbon pile control element. A spring associated with the solenoid provides a mechanical reference. With proper connections, the loop is degenerative. T_1 is used to provide derivative negative feedback around the solenoid and carbon pile element to obtain optimum dynamic performance of this electromechanical device.

One of the most troublesome features of a carbon pile regulator is the change in characteristics of the carbon pile with temperature, time, and use. This affects the dynamic and static stability and necessitates frequent checking and adjustment of the regulator. Chapter 12, Vol. 17, contains a complete discussion of the critical adjustment procedure employed with this type regulator.

The carbon pile is separately mounted on shock mounts, because vibration modulates the resistance at the vibration frequency. High-frequency vibration is not serious, since the output of this type of inverter has a poor wave shape anyway and a few per cent increase in harmonic content makes little difference. However, low-frequency vibration may be serious.

16·5. Regulated A-c Oscillators.—There are many circuit applications that require voltage-regulated a-c power at a frequency not readily available. Electronic computers of all sorts are an important class of circuits requiring such supplies. If the carrier frequency is low or intermediate, it can usually be generated most efficiently by a Wien bridge oscillator. Some electronic computers require more accurately stabilized frequency and less accurately stabilized output. For these applications Wien bridge circuits are generally impractical. Crystal or tuning fork circuits are the most efficient when high-precision frequency control is required.

Constant-voltage Wien Bridge Oscillators.—The circuit shown in Fig. 16·24 was designed at Bell Telephone Laboratories with the participation of Radiation Laboratory engineers for use with an airborne computer requiring accurately regulated voltage insensitive to element aging and temperature. The output of this circuit is 120 volts, 350 cps, working into a 4000-ohm load. For a period of one week, for a temperature range of $-50°$ to $70°C$, and for a 10-per cent change in load current, the amplitude shifts by less than 1 per cent. The dynamic response has a time constant of $\frac{1}{2}$ sec. The frequency, depending on R_1, R_2, C_1, and C_2, is constant to approximately 1 per cent. Harmonic content is less than 0.1 per cent when a WE1C Thermistor is used; other Thermistors may introduce distortion.

The tuned output transformer provides an efficient coupling network between the output amplifier and the load. A balanced center tap on the output transformer ensures two symmetrical signals of opposite sign for the bridge. The gain of the network at this frequency is controlled

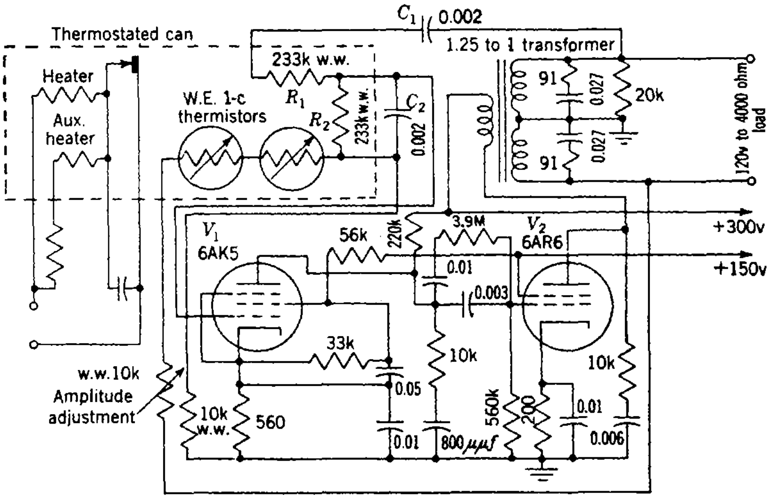


FIG. 16·24.—Precision 350-cps oscillator.

by two Thermistors in an arm of the bridge. As the frequency is nearly independent of the Thermistor characteristics, and since the frequency-setting elements are not sensitive to power input, the amplitude controlling is done by the Thermistors alone. The Thermistors decrease resistance as the power applied increases; this reduces the bridge gain, and a stable amplitude condition is established when the loop gain is unity.

Constant-frequency Feedback Oscillators.—Figure 16·25 is a circuit diagram of a 120-cps oscillator with frequency stability better than 0.005 per cent for a temperature range of $-50°$ to $+70°$C and for a period of one year.[1] The circuit was designed to provide an accurate time base for an airborne computer. Vibration as encountered in aircraft has no effect on the frequency stability for either a short time, cycle to cycle, or long time, over a period of minutes.

[1] This oscillator was designed by American Time Products Corp., New York.

This tuning fork was a compromise design in which weight and space were important and accuracy required was not nearly so high as it is possible to attain. Tuning forks can be obtained in a frequency range of 240 to 1000 cps, temperature compensated to 1 part in a million over a temperature range of 15° to 75°C.

The oscillator is made up of a two-stage resistance-coupled amplifier and two feedback loops. The tuning-fork unit provides the positive feedback and mechanically resonant frequency-determining circuit. The amplitude stabilizing negative feedback loop is made up of a Varistor[1]

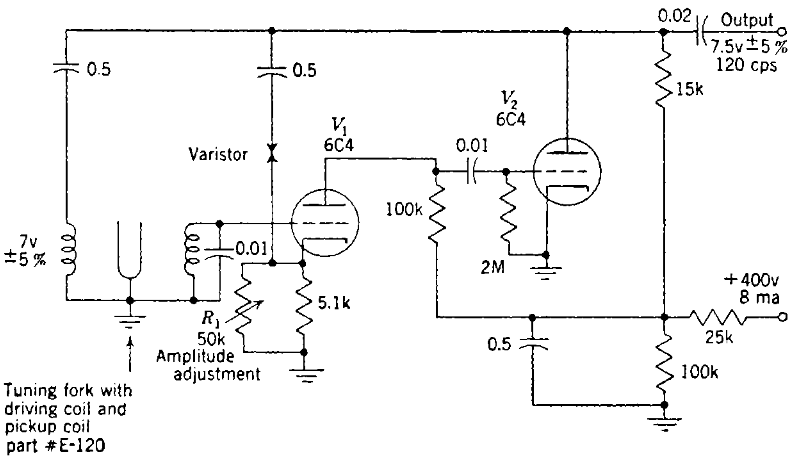


FIG. 16·25.—Constant-frequency oscillator using tuning fork.

and a resistance in the cathode of V_1. The tuning-fork assembly, E-120, consists of a bimetallic temperature-compensated tuning fork, a drive coil, and a pickup coil, all of which are enclosed in a pressurized, hermetically sealed metal container.

The negative feedback loop is adjusted by R_1 to provide an output of 7.5 volts rms, ±5 per cent, and because of the nonlinearity of the Varistor element the output is essentially constant. The Varistor distorts the output slightly, but this distortion was desirable in the original application of this circuit, as it provided better control of the thyratron inverter tubes that followed.

16·6. Precision D-c Voltage Supplies.—Many electronic instruments require voltage supplies that are stable for a considerable time to better than one part in a thousand and recover from load transients with a time constant of the order of $\frac{1}{1000}$ sec. The circuits given in this

[1] Varistor is the Western Electric trade name for a nonlinear resistance material made of silicon carbide. Similar material is called Thyrite by General Electric.

section all contain elements that are rugged enough to stand commercial or military use. For laboratory purposes, voltage supplies need not be so rugged. Section 16·7 contains many circuits that have found wide application in laboratory use.

Figures 16·26 to 16·28 are circuit diagrams of the conventional types of regulated d-c voltage supplies. The component values are marked on the diagrams, and the average performance characteristics are given in Table 16·4.

A Precision Voltage Supply.—The following precision voltage supply is rated 105 ma at +200 volts and 105 ma at −200 volts with less than 0.1 per cent variation in either voltage for a 10 per cent change in line voltage, a 30 per cent change in load, and under conditions of vibration, shock, and temperature encountered on combat ships. Figure 16·29 is the top view of the chassis.

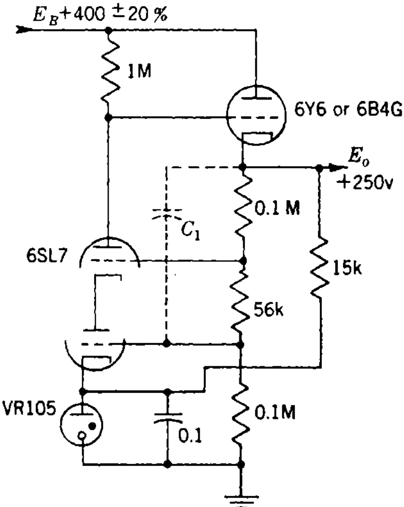


FIG. 16·26.—Cascode regulator.

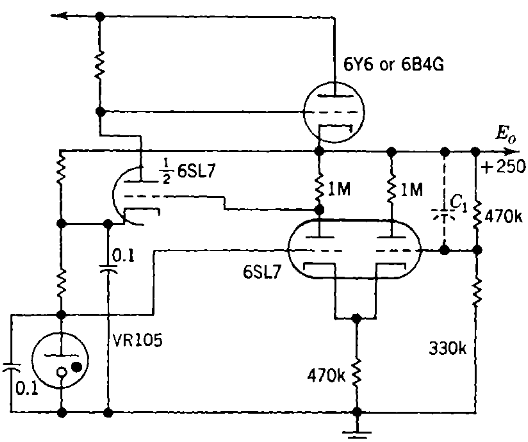


FIG. 16·27.—Balanced comparison circuit.

A vibration test[1] was run on a model of this supply with amplitude 0.36 in. The d-c output deviation was under 0.1 per cent except at

[1] RL Report No. 1217. Vibration tests of "No. 97 Precision Voltage Supply."

Table 16·4.—Estimated Performance of Several Regulator Circuits

Fig. No.	D-c regulation, %*	D-c impedance, ohms	120-cycle regulation, %†	120-cycle impedance, ohms	Output current range, ma to load	Output voltage change for 10% change of heater voltage of equating tube	Output voltage stability for 1 week V-R reference	Output voltage stability for 1 week battery reference
16·3 VR-105	0.10	40	May be two to three times d-c value	May be two to three times d-c value	0–30		±0.5 in 105.0	
16·10 1/2 6SL7	1.25	17	0.60	8.00	0–60	1.00	2.0	1.00
16·10 6SJ7 E_{g2} 25 v	0.75	4	0.35	2.00	0–60		2.0	1.00
16·26 Cascode 6SL7	0.20	0.9	0.10	0.45	0–60	1.00	2.0	1.00
16·27 Balanced equating circuit	0.05	0.8	0.02	0.40	0–60	0.15	1.5	0.25
16·27 Balanced equating circuit 4-26A7 tubes in place of 1-6B4-G§	0.04	0.7	0.18	0.35	0–60	0.15	1.5	0.25

* Regulation percentage change of d-c output for a 20 per cent change in line voltage. Load current constant.

† 20 per cent 120-cycle ripple would be reduced to this percentage in the output.

§ This variation to use 4-26A7 tubes results in no gain in performance.

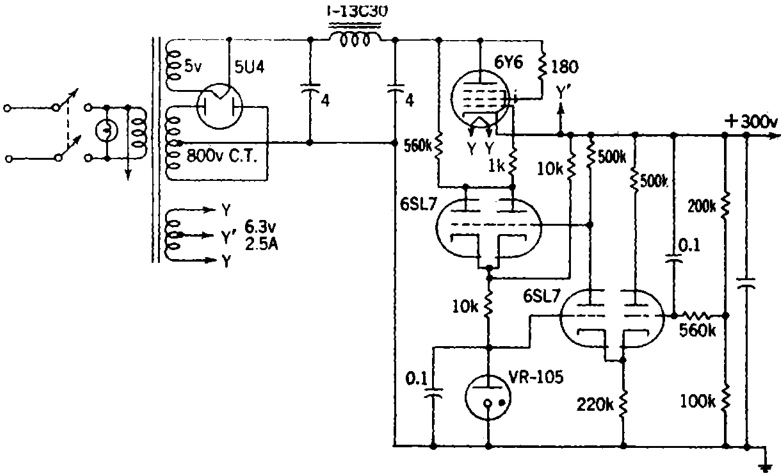


FIG. 16·28.—Balanced comparison circuit.

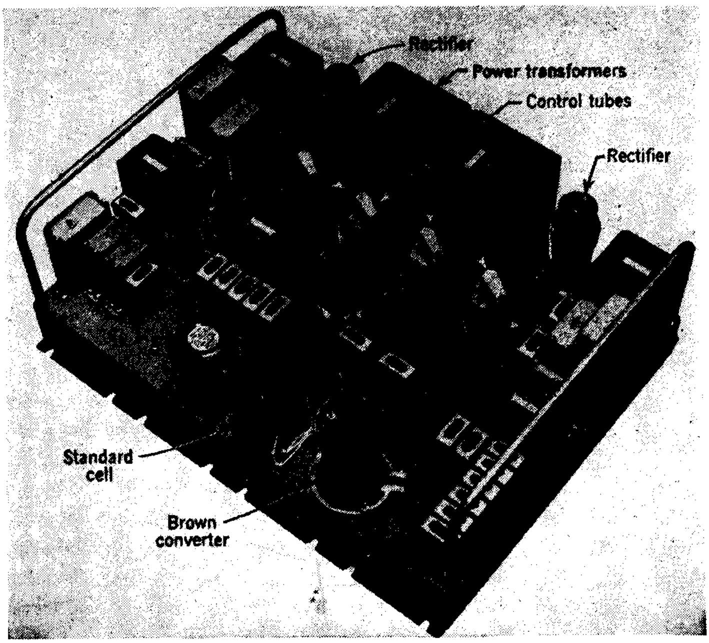


FIG. 16·29.—Precision voltage supply.

17 cycles, at which frequency the chassis resonated violently and the error was 1 per cent. The a-c ripple was 25 to 60 mv, being a maximum at the resonance of the chassis.

Figure 16·30 is a block diagram of the −200-volt supply showing the two loops that together achieve the desired performance. Loop *a* is a conventional regulator using a VR-tube reference. With the phase detector short-circuited, this loop is adjusted to give approximately −200 volts. With loop *b* in operation and the phase detector operating, the sampling circuit *b* is adjusted to give precisely −200 volts output.

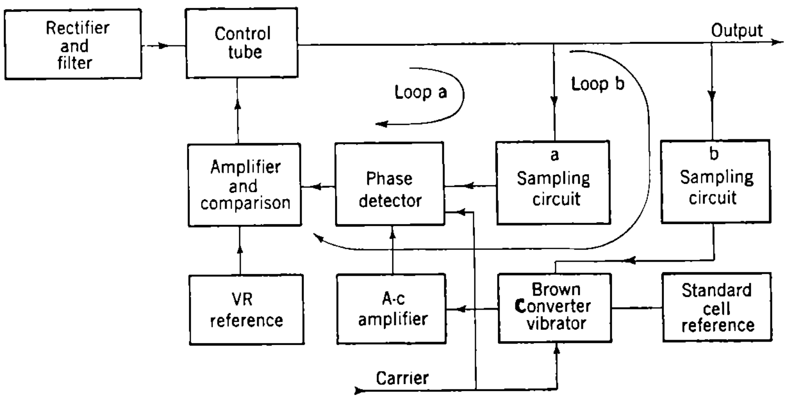


FIG. 16·30.—Block diagram of d-c regulator with standard cell reference.

Any deviation in the output as seen at the contacts of the vibrator produces an a-c signal representing the error in magnitude and phase. This signal is amplified and detected by the phase detector and corrects loop *a*. A 60-cycle carrier is applied to the vibrator and to the phase detector to produce the desired switching action. Loop *b* has a bandwidth of about 5 to 10 cycles owing to the use of a 60-cycle carrier and to the use of filters to smooth the phase detector output. Loop *b* can correct only for relatively slow changes, while loop *a* can correct for changes of several thousand cycles per second. The use of an output capacitor reduces the impedance at high frequencies.

Figure 16·31 is a schematic diagram of this supply. The reference element is a standard cell. The vibrator compares the sum of the sampled voltage and reference emf with ground. R_1 and R_2 are used to pass a portion of the plate current, thus reducing the required current capacity of the series control tubes. This is a practical way to reduce the number of tubes if the current range is not too great. R_3 limits the current through the standard cell, and relay K_1 is not closed until the negative supply is warmed up. This also keeps the current through the cell low.

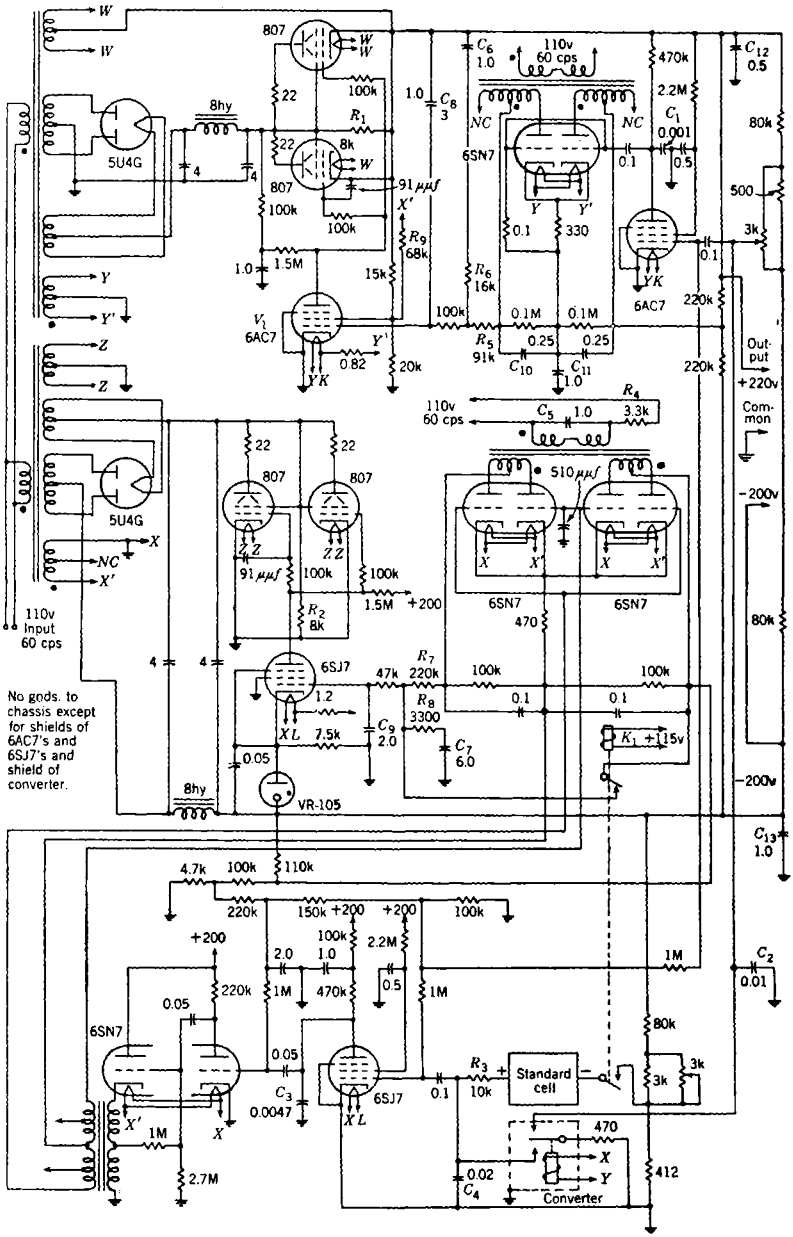


FIG. 16·31.—Precision d-c voltage supply.

The Brown Converter is actuated by 9.5 volts alternating current instead of 6.3 volts (for which it is designed) to reduce effects of vibration on the accuracy of the comparing operations.

C_1, C_2, C_4, and C_3 reduce high-frequency pickup, remove any spikes from the square wave, and prevent high-frequency oscillation of the amplifiers. R_4 and C_5 shift the phase of the 60-cycle carrier to compensate for the phase shift in the vibrator and a-c amplifier.

The integrating circuits, made up of R_5, R_6, C_6, and R_7, R_8, C_7, act to stabilize loop b for each regulator. C_9 and C_8 are a part of loop a for

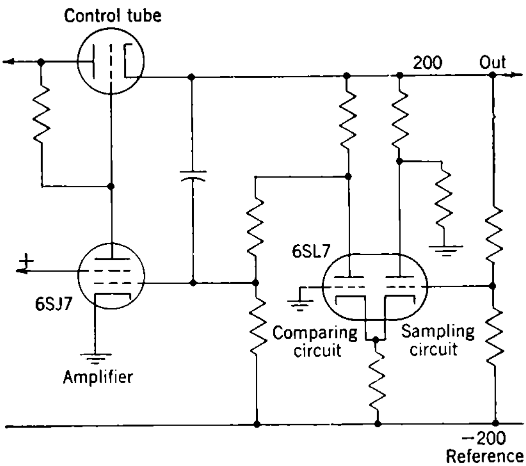


FIG. 16·32.—Alternate circuit for +200-volt regulator.

each regulator increasing the middle-frequency gain of these loops. The d-c gain from vibrator to phase detector output is about 22,000 for the negative regulator. Gain for the positive regulator can be one hundred times less, since the reference voltage is effectively one hundred times greater.

The phase detector for the positive supply has a 60-cycle ripple component which is not entirely removed by filter C_{10} and C_{11}. R_9 applies a 60-cycle signal to the screen of V_1, the magnitude and phase of which is of such a nature as to cancel the ripple from the detector.

C_{12} and C_{13} reduce the impedance of the regulators at the higher frequencies.

Alternative Circuit Arrangements.—The use of a vibrator comparison circuit, a-c amplifier, and phase detector would not be justified if a reference voltage of -200 were available. An alternative circuit is shown in Fig. 16·32. All data available indicate that a 6SL7 comparing tube in a balanced circuit can be depended on to stay within ± 50 mv for a considerable period of time (see Section 15·4). With a sampling circuit having an attenuation of 2.0, the required stability of the comparing circuit would need to be no better than 100 mv for 0.1 per cent stability error from this source.

The capacitor minimizes any stray pickup in the 6SL7 comparison circuit and prevents ripple in the −200 volts supply from affecting the

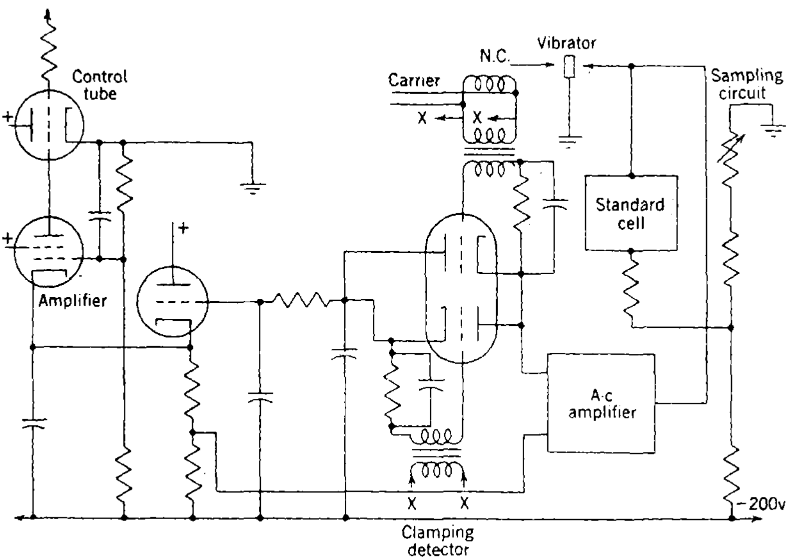


FIG. 16·33.—Alternative 200-volt regulator for the circuit of Fig. 16.31.

+200-volt regulator. It also divides the regulator into two loops, one around the 6SJ7 and control tube, which is effective for frequencies above a few cycles per second, and a loop through the sampling circuit, 6SL7 comparison circuit, 6SJ7 amplifier, and control tube in which only slow changes at a rate of less than a few cycles can occur.

The use of VR tubes in these regulators is not justified, as VR tubes may introduce some roughness of small amplitude at a rate higher than the precision correcting circuit can function. The VR tube and phase detector in Fig. 16·31 might be replaced by the circuit shown in Fig. 16·33.

If this arrangement were used with a 400-cycle vibrator, the response time could be made high enough around the precision loop practically to eliminate all effects of vibration. The clamping detector shown is just one of a number of detectors that would be satisfactory. This circuit arrangement has the advantages of fewer components and fewer adjustments.

16·7. Laboratory Regulated D-c Supplies.—The d-c supplies described in this section are useful in the development of electronic instruments and in other experimental work.

Example A.—This is a commercially available unit built by the Oregon Electronics Manufacturing Company. The outputs provided are unregulated 400 volts and adjustable-regulated 0 to 300 volts at a total of 200 ma (150 ma is limit on the regulated output); adjustable negative 0 to 150 volts, VR-stabilized at a maximum of 2 ma; and an a-c output of 6.3 volts at 5 amp.

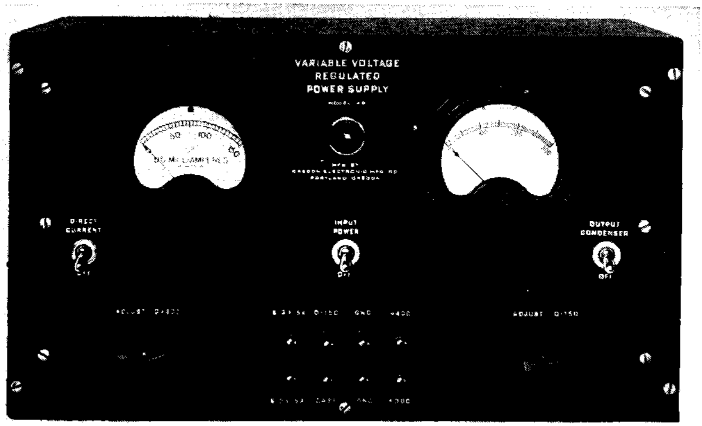


FIG. 16·34.—Regulated power supply Example *A*.

The construction of this unit is shown by Figs. 16·34 to 16·36 which are respectively front, top interior, and bottom interior views. Figure 16.37 is the schematic circuit diagram of this supply. The output impedance and regulation of the negative supply of this unit are approximately 30,000 ohms and 0.15 per cent for a 10 per cent line change. It will supply only 2 ma. The regulated 0- to 300-volt output when set at 250 volts has an impedance of 15 ohms for direct current. Figure 16·38 shows how impedance varies with frequency. The regulation is rather poor, being approximately 1 per cent for 10 per cent line change. The apparent discrepancy between the low output impedance between 20 and 20,000 cycles (Fig. 16·38) and 15 ohms at direct current is due to the presence of the 0.1 μf condenser shunting R_1. Also the unregulated source for the VR reference is not affected by the alternating current, and C_3 decreases the impedance for alternating current. These three factors make a marked difference in the output impedance of this regulator. The increase at 100 kc is due to phase shift in the negative feedback loop. The relatively small effect of the 4 μf condenser C_3 on the imped-

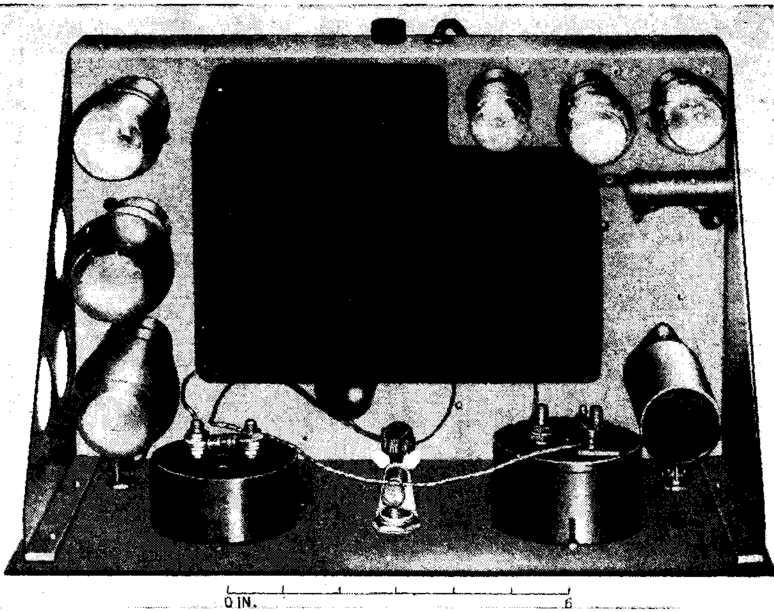


FIG. 16·35.—Top interior of regulated power supply, Example A.

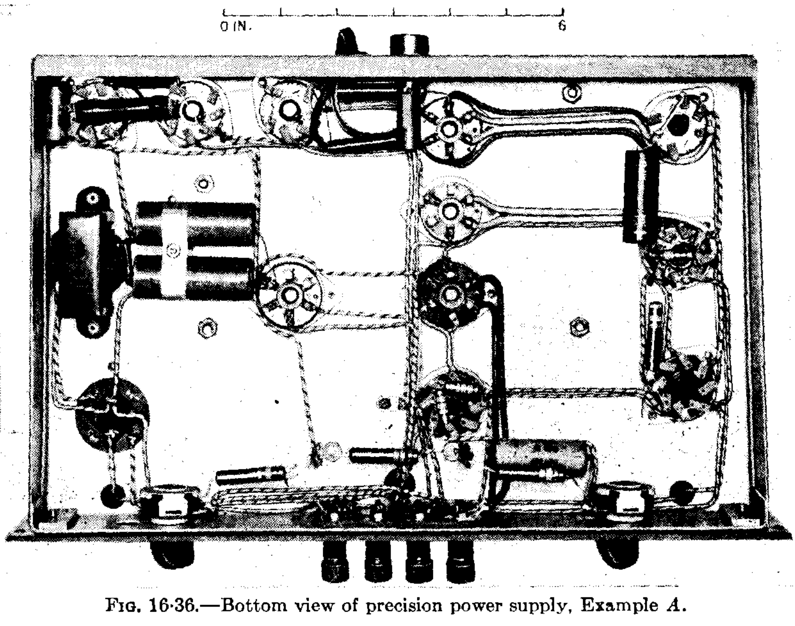
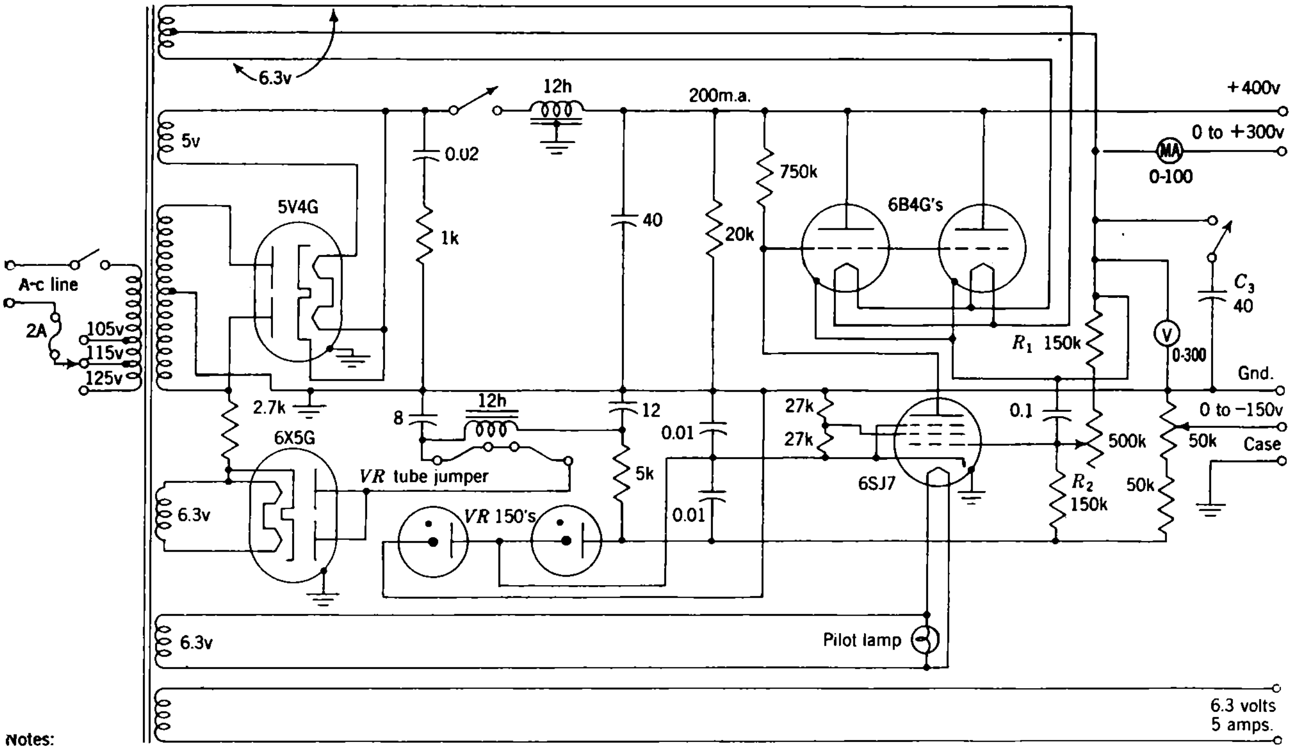


FIG. 16·36.—Bottom view of precision power supply, Example A.



Notes:
1. The 150k resistors used as R_1 and R_2 should he individually measured and the larger of the two used as R_2

FIG. 16·37.—Variable voltage regulated power supply.

ance at high frequencies should be noted in Fig. 16·38. There are a number of reasons why the performance of this regulator is poor compared with other circuits in this section. One factor is that the screen voltage on the 6SJ7 amplifier is higher than it should be; reduction by 25 or 30 volts would improve operation materially. The use of the unregulated supply to operate the reference VR tubes is another factor that impairs performance. Despite these performance limitations, this unit is valuable for laboratory use.

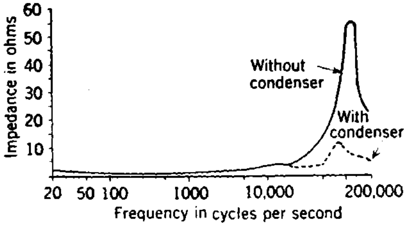


FIG. 16·38.—Output impedance vs. frequency characteristics of Oregon power supply.

Example B.—This regulated power supply (Fig. 16·40–44) was widely useds at the Radiation Laboratory for experimental and developmental work. The outputs provided are a total of 225 ma from the unregulated 450-volt bus and regulated 250-

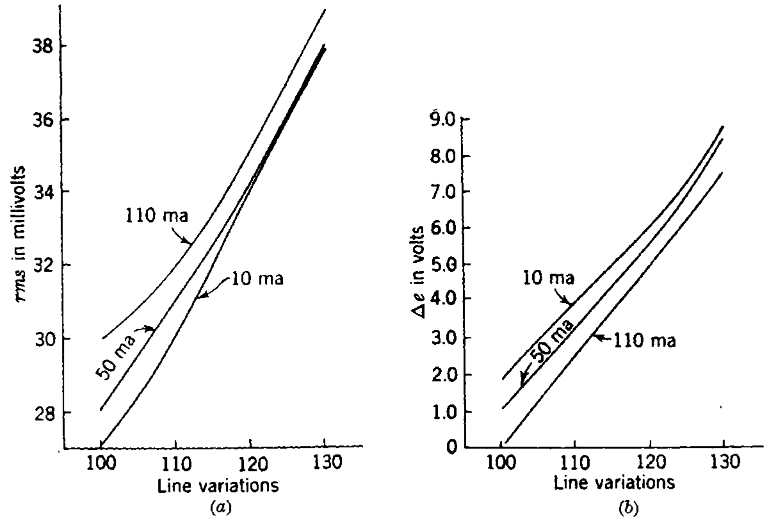


FIG. 16·39.—Characteristics of Oregon Electronics supply. (a) Output ripple; (b) change in output voltage.

volt bus, 30 ma from the VR-regulated −105 volt bus, and 8 amp alternating current at 6.3 volts. An unregulated supply at −1700 volts 2 ma, and 6.3 volts at 1 amp, are provided for a cathoderay tube.

The performance of the negative VR-stabilized supply is: output impedance; 50 to 100 ohms, and regulation, 0.15 per cent for a 10 per cent line voltage change. Figure 16·40 shows the performance of the

250-volt supply for line and load changes. The output impedance for direct current is 20 ohms, and the regulation is approximately 0.1 per cent for a 10 per cent change in line voltage.

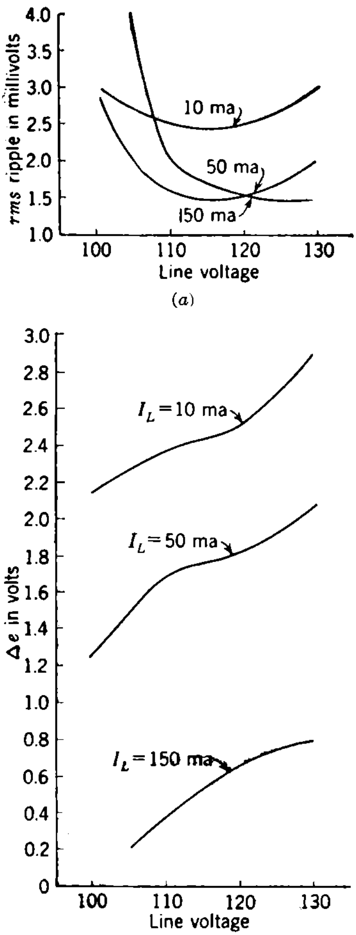


Fig. 16·40.—Characteristics of laboratory power supply. (*a*) Output ripple; (*b*) output voltage.

The compensating network R_1 and R_2 modifies the characteristics of this regulator materially. In this case R_2 is adjusted to minimize the effect on the output voltage of slow changes in line voltage. This compensation has to overcorrect for regulation to correct for the change in heater voltage on the comparing tube. The result is a momentary change of 0.5 per cent in output voltage for a 10 per cent change in line voltage settling back to the values given in Fig. 16·40 with a time constant of a few seconds.

The variation in output impedance with frequency (Fig. 16·44) from 20 ohms for direct current dropping to less than 1 ohm at 500 cycles and rising to above 30 ohms at 125 kc is typical of this circuit. The rise in output impedance above 25 kc is due to loss in the feedback gain and phase shift in the negative feedback loop.

The compensation network, a fairly well smoothed unregulated supply, and C_1 are responsible for the low output ripple (see Fig. 16·40).

Example C.—This is a modification of Example *B* in which the newly developed 6AS7G tube replaces 3-6B4 tubes as the control element. The output was made +200 volts at 100 to 200 ma. The 200-ma rating, low plate-cathode potential for rated current, and 300-volt heater-to-cathode rating of the 6AS7G are advantages; however, the low μ (2.1) is a limitation in the use of this tube, as the grid swing required may necessitate a more complicated amplifier.

The unregulated supply variations were reduced somewhat by using choke input. The allowable current variation is less owing to the inability

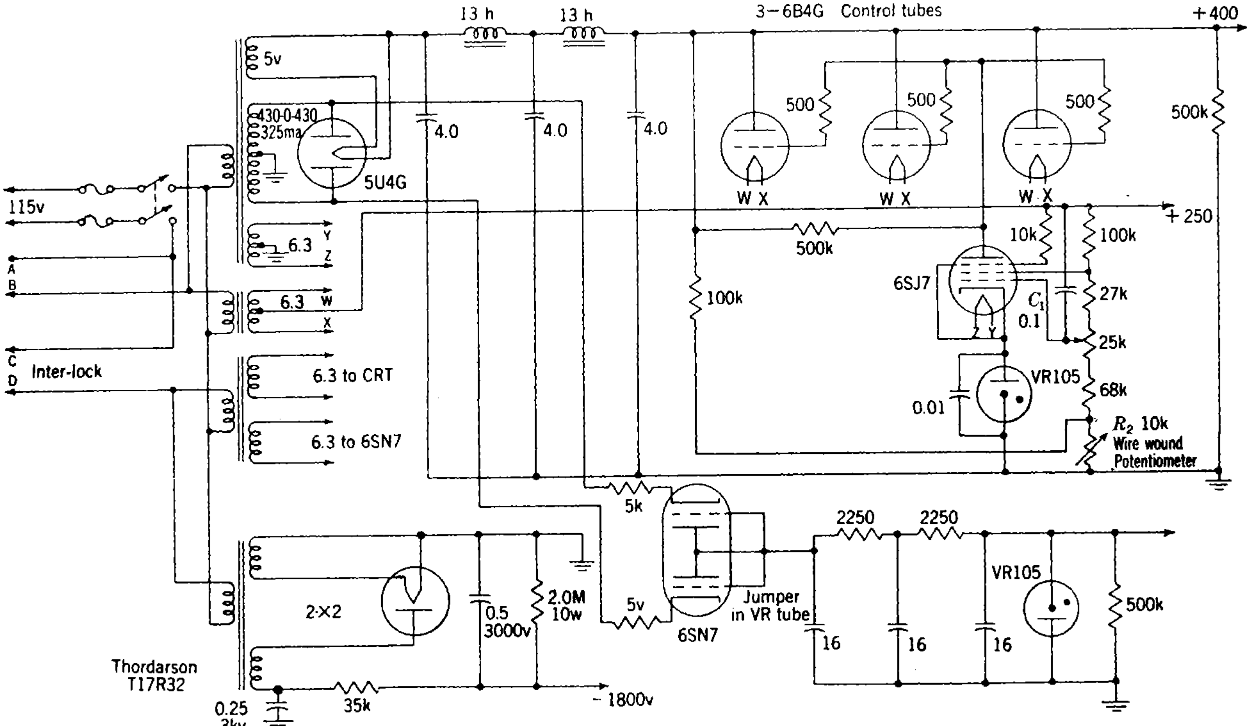


Fig. 16·41 —Schematic of laboratory power supply, Example *B*.

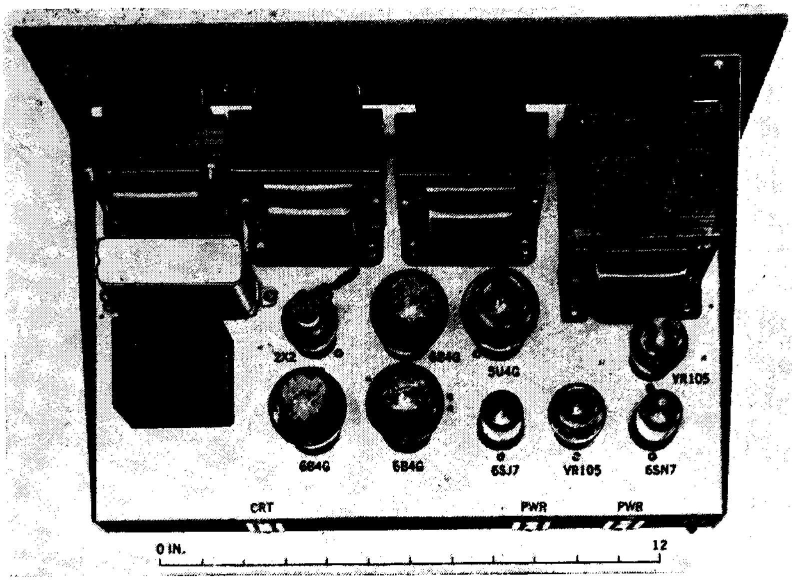


FIG. 16 42.– Laboratory power supply, top interior view.

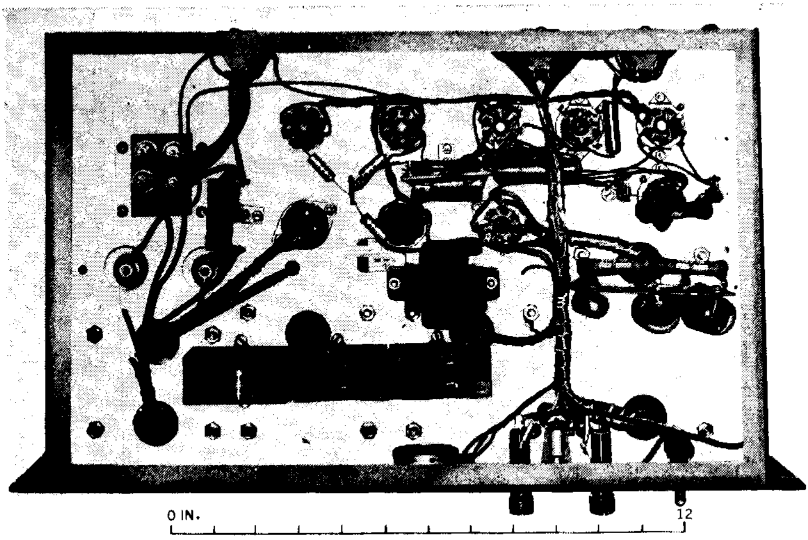


FIG. 16·43.—Laboratory power supply, bottom view.

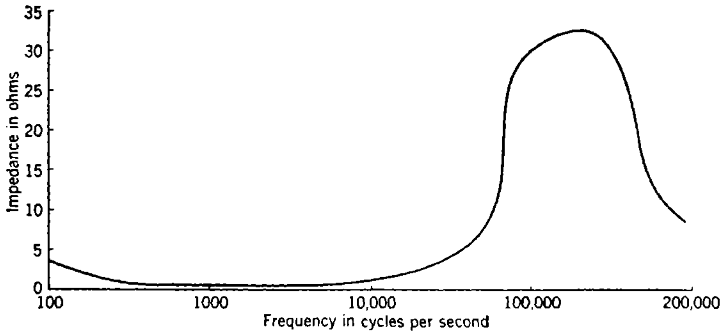


FIG. 16·44.—Output impedance vs. frequency for laboratory regulated power supply.

of providing adequate grid voltage variation to the control tube with the 6SJ7 amplifier connected in the usual manner. The operation for line voltage of 100 to 130 volts and load current 100 to 20 ma is shown in Fig. 16·45. Figure 16·46 indicates the variation of impedance with frequency. The regulation is 0.14 per cent for 10 per cent change in line voltage, and the output impedance for direct current is approximately 15 ohms for operation at the higher load currents. The lack of linearity at lower currents is due to the lower μ of the 6AS7-G requiring more voltage swing out of the 6SJ7. This is a serious limitation. The comments on the compensation circuit of Example *B* are the same for this example. The larger change in ripple for line and load variations results from more nonlinear operation reducing the effectiveness of the compensation.

Example D.—In an effort to obtain the smoothest possible output, very low ripple, very low output impedance, and very good regulation the circuit of Fig. 16·47 was developed.[1] This regulated supply has an output of 250 volts at 0

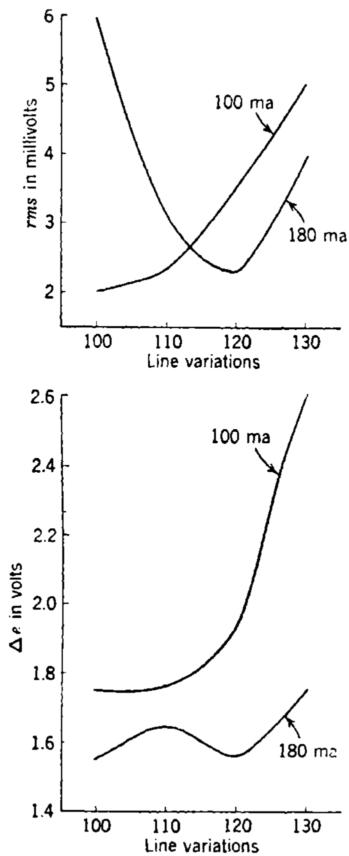


FIG. 16·45.—Characteristics of regulator *C*.

[1] J. L. Lawson, "Notes on Design and Construction of Regulated Power Supplies," RL Report No. 44, Feb. 26, 1945.

to 350 ma with ripple of less than 0.3 mv rms, an output potential change of about 15 mv for the rated current range and less than 5 mv for 105 to 125 volts input. The source impedance for direct current is less than 0.2 ohms, and the regulation factor is better than 0.002 per cent for a 10 per cent line voltage change. These rather fantastic figures for regulation

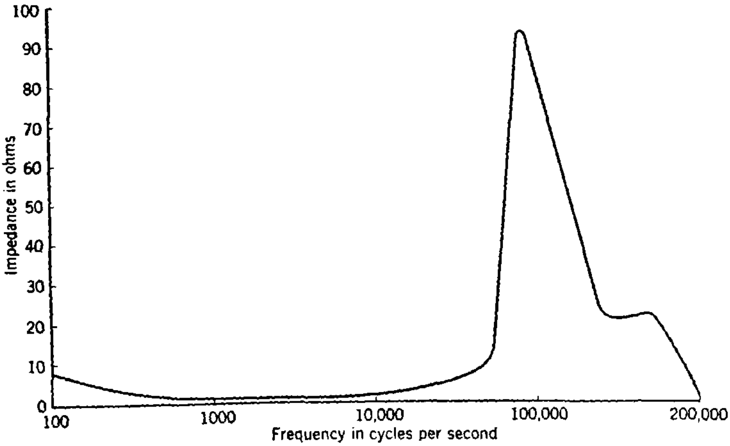


FIG. 16·46.—Output impedance vs. frequency for regulator *C*.

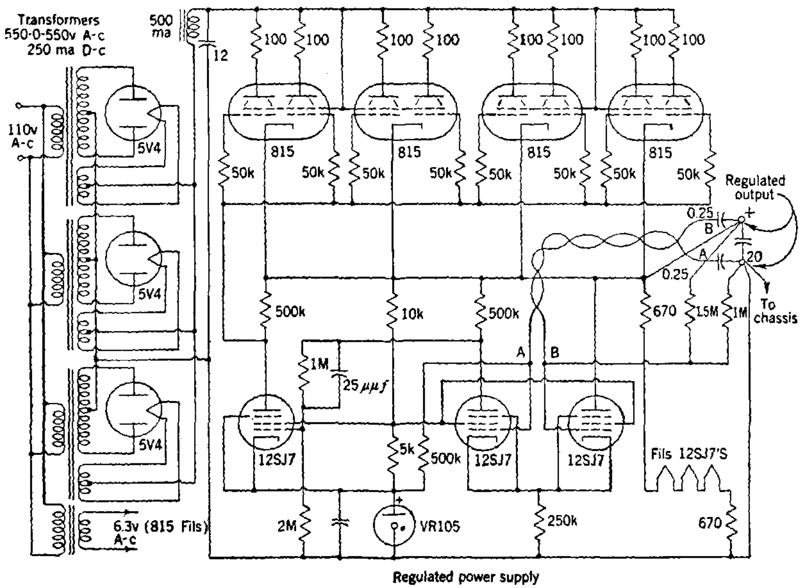


FIG. 16·47.—Regulated power supply.

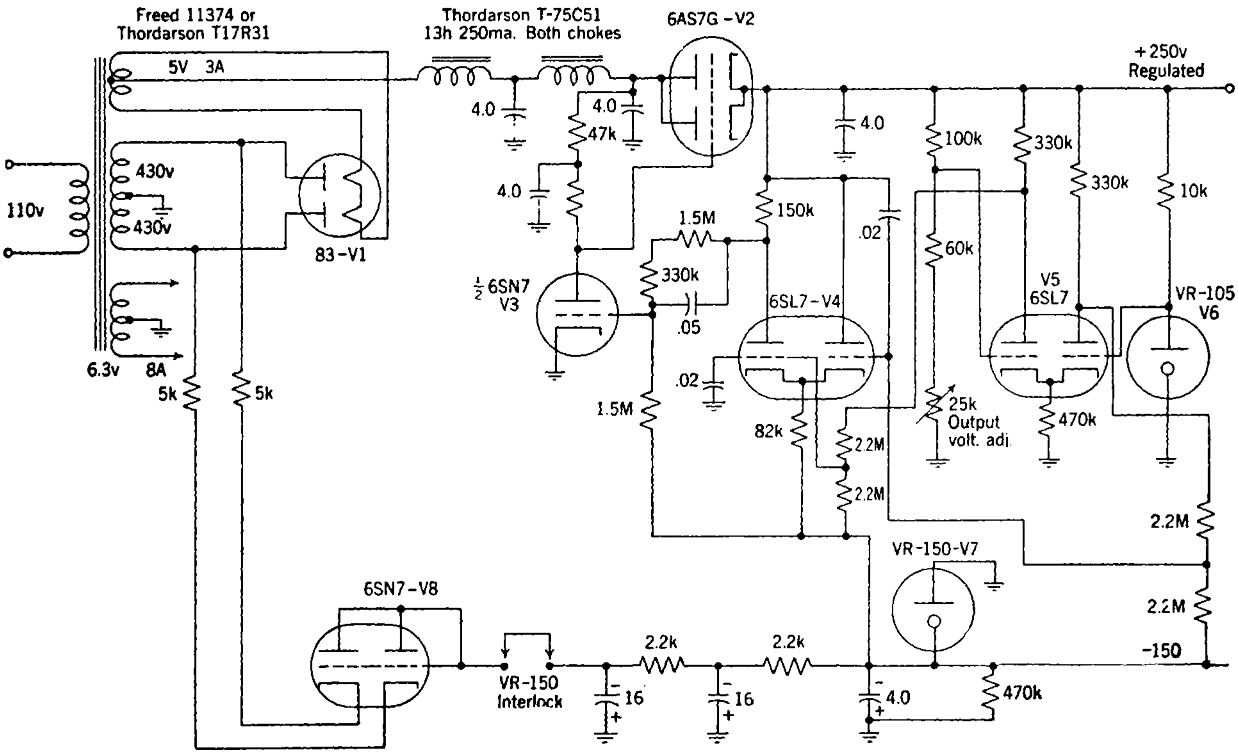


FIG. 16·48.—Direct-current regulator with balnaced comparison circuit.

and source impedance as well as the low ripple are achieved by a number of expedients.

The circuit has a balanced pentode comparing circuit with an additional stage of amplification using another pentode and four 815 dual beam power tubes, triode-connected in parallel, for the control element. The impedance of the unregulated supply is kept low by the use of three power transformers and rectifiers in parallel. The balanced comparison circuit has the 150-ma heaters connected to the regulated output so that no ripple is introduced through the cathodes and no unbalance occurs due to heater voltage changes.

The effects of common resistance paths and electrostatic and electromagnetic coupling causing spurious signals between the output and the comparison tubes are minimized by the use of *RC*-filters at the grids of the comparison amplifier and by careful attention to the wiring and arrangement of components.

The static or long-time stability is limited primarily by the VR tube and the effects of leakage in the high-impedance *RC*-filters in each comparison tube grid. With series resistance of 500 k, 50-megohms leakage resistance causes 1 per cent error. Leakage effects tend to vary widely with temperature and time, so for long-time stability of the order of 0.5 per cent both a different sampling circuit and reference would be necessary.

Example E.—This regulator (Fig. 16·48) is designed to have good long-time stability, very low output impedance, and very good regulation. The use of a tube comparison circuit and VR tube for reference are limitations to long-time stability. This circuit is well adapted to the use of a dry battery reference voltage in place of the VR tube.

The supply provides up to 200 ma at 250 volts with a regulation of 0.045 per cent for a 10 per cent line voltage change and has an output impedance of 0.16 ohm for direct current. Figures 16·49 and 16·50 are typical performance curves for line and load variations. The −150-volt bus at about 20 ma has the usual characteristics of VR-regulated supplies: about 60 ohms output impedance and regulation of about 0.15 per cent for a 10 per cent line voltage change. The effect of a fraction of a volt fluctuation of the −150-volt bus on the regulated 250-volt output is practically zero due to the amplifier design.

The regulator as shown in Fig. 16·48 stays within $\frac{3}{4}$ of 1 per cent of the original voltage over a period of one week. The variation is determined mainly by drift of the VR tube.

For the comparison circuit a 6SL7 balanced amplifier operated at low plate current is used. A wire-wound resistance sampling circuit is used to minimize drift. The mercury vapor rectifier and choke input filter make the unregulated supply have a low impedance. A 6AS7-G control tube with a three-stage amplifier aids in realizing low output impedance.

The amplifier V_3 which provides the grid voltage for the 6AS7-G tube has a relatively low value of plate-load resistor to improve the response at higher frequencies. Additional filtering by an *RC* combination is required for the plate supply of the V_3 amplifier to minimize the effect on the output voltage of 1-volt ripple of the unregulated supply. V_4 functions as a differential to single-ended comparing amplifier for the intermediate and high frequencies.

16·8. Miscellaneous D-c Regulators.—There are several regulating circuits that have found general application but do not fall into any of

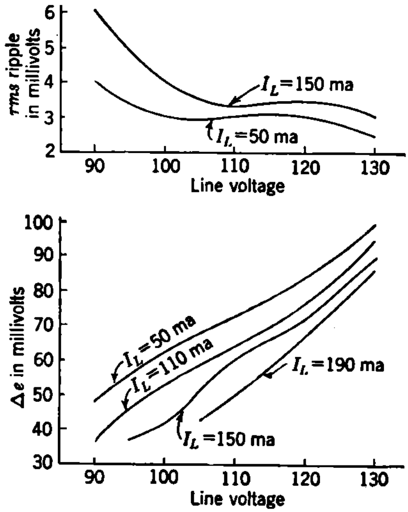


FIG. 16·49.—Characteristics of regulator with balanced comparison circuit.

the classifications so far discussed in this chapter. Perhaps the most important of these is the use of VR tubes in simple regulator circuits. Another class that has not been discussed is the regulation of high-voltage low-current supplies for application to cathode-ray tubes, electron multipliers, etc. Some typical circuits for these applications are given here. For more detailed information on the regulation of cathode-ray tube supplies refer to Vol. 22.

Simple VR-tube Regulators.—As simple regulators, VR tubes (and other glow tubes) are used in series resistance, shunt tube circuits. When a VR tube is used as a simple regulator, the flatness of the volt-ampere characteristic and the low effective impedance are probably the most important characteristics. R_g varies considerably from tube to tube and from manufacturer to manufacturer. Figures 16·4 and 16·5, in Sec. 16·2, showing the variation in R_g with frequency and current, are typical of VR tubes. Other glow tubes have more variable characteristics. Figure 16·51*a* and *b* show the operating conditions for two of the most popular VR tubes. These curves together with the data in Table 15·1 are all that are needed to design a simple regulator using VR tubes and to predict its performance. If the increase in internal impedance with fluctuation frequency is undesirable, it can be reduced by connecting a condenser across the VR tube.

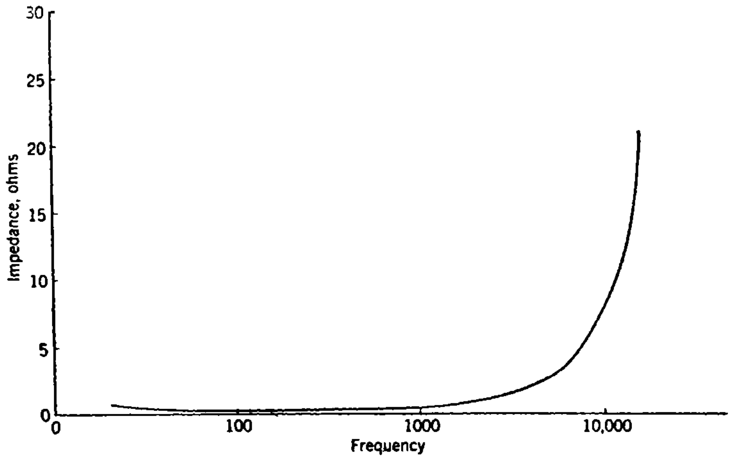


FIG. 16·50.—Characteristics of regulator *E*.

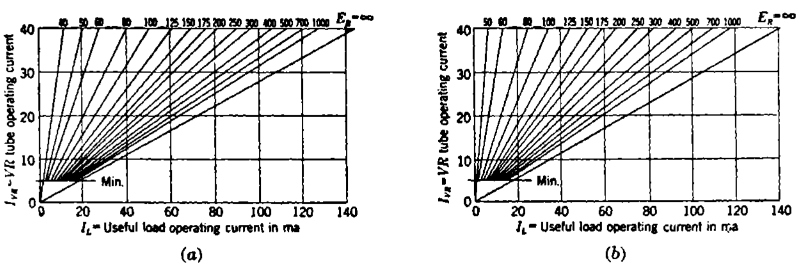


FIG. 16·51.—Operating conditions for certain striking of tubes. (*a*), VR-105; (*b*), VR-150. E_r is the drop between source and regulated output.

High-voltage Regulators.—The use of series control tubes may be impractical when the voltage to be regulated is high. However, if the current drain is very low, other forms of high-voltage regulated supplies are practical. In this case the load impedance may be very high, and the internal impedance of the regulator will not greatly affect operation. This is the case for supplies designed to operate cathode-ray tubes. It is, however, necessary to regulate against line voltage changes, since the focusing, intensity, and deflection sensitivity of cathode-ray tubes are sensitive to the anode voltages.

Figure 16·52 is typical of a class of high-voltage regulated power supplies. The high voltage is obtained by rectifying the stepped-up output of a stabilized audio oscillator. The output impedance of such

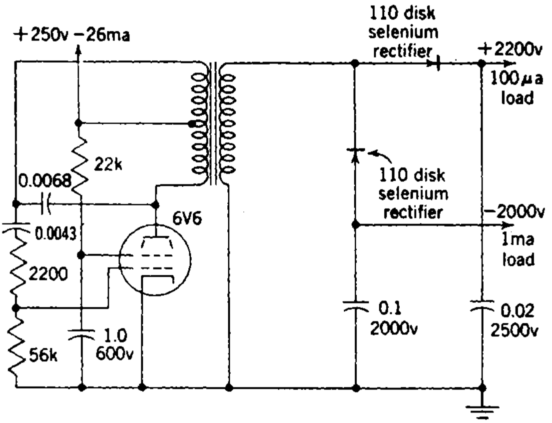


Fig. 16·52.—Schematic of 2 kv a-f high-voltage supply.

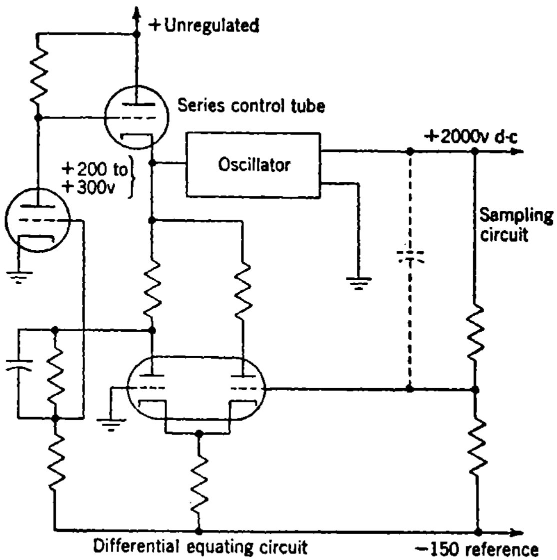


Fig. 16·53.—Oscillator high-voltage supply with d-c regulator.

a circuit is quite high, but this is not too troublesome if the load is a cathode-ray tube with a constant impedance bleeder. It is not suitable if large leakage currents due to moisture condensation, etc., are encountered. By proper choice of elements the oscillator output is made independent of small variations in the input voltage. The use of selenium rectifiers eliminates the necessity of having vacuum-tube rectifiers. Except for the problem of heater supply, vacuum-tube rectifiers are preferable. The 8016 is the tube most frequently used, as it is specially constructed for this type of application, has a $\frac{1}{4}$-watt filament, and is designed to be operated at audio and radio frequencies.

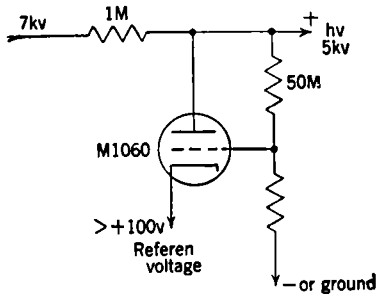


FIG. 16·54.—Degenerative shunt high-voltage regulator.

Figure 16·53 is an extension of the circuit in Fig. 16·52 to improve the regulation and output impedance. It is a completely degenerative regulator and hence has better performance than the circuit of Fig. 16·52. Since it requires more tubes and a lower impedance divider (for the sampling circuit), the circuit of Fig. 16·53 has much lower efficiency.

Figure 16·54 shows the application of a special shunt tube to regulate high voltage. The reference element must have several hundred volts output but can have fairly high impedance. This circuit is a special application of the transconductance bridge (see Fig. 16·8), and the conditions for balance are given in Sec. 16·2.

PART IV
PULSE TEST EQUIPMENT

CHAPTER 17

DEVELOPMENT AND DESIGN

BY H. J. REED, JR., AND R. W. LEE

INTRODUCTION

17·1. Scope of Part IV.—The term "test equipment" has been applied generally to all of the auxiliary apparatus used in the design and maintenance of electronic equipment. This includes general-purpose laboratory instruments, permanently constructed apparatus for special applications, and "breadboard" units which are built for a limited number of special measurements in the laboratory. Radar test equipment may logically be classified in four fairly well defined groups: (1) units used for measurements at microwave frequencies, (2) those used at intermediate frequencies (i-f), (3) those used at video frequencies, and (4) those used in the d-c, audio, and supersonic frequency range.

In Part IV, equipment designated as video or pulse circuit test equipment will receive the most emphasis. The upper frequency limit involved may be as high as 30 mc, for pulses containing such frequency components are sometimes employed as signals or for synchronization. Test equipment in this group includes oscilloscopes with synchronized sweeps, synchronizers that supply trigger pulses and beam deflection voltages for an external cathode-ray tube, pulse and marker generators, amplifiers with wide bandpass characteristics, and apparatus for the precise measurement of small time intervals. It is noted that the circuits involved are very similar to those in the ranging and indicating sections of radar systems.

The present chapter will deal with some of the more important aspects of test equipment development and design. It is hoped that the material covered in this chapter will be of assistance to the general reader in following some of the necessarily terse descriptions of representative equipments in the following chapter. An object of the authors has been to provide a background which will be especially helpful to the reader interested in adapting to his needs portions of the many circuits of Chap. 18. A search for specific circuits and a general understanding of over-all equipments is greatly facilitated by an understanding of the fundamental circuit types from which the designers have put together the pulse test equipments listed in Chap. 18, and of some of the military requirements which greatly influenced these designs. It is with this purpose in mind that the material of Chap. 17 is presented.

At the start of the war period, test equipment applicable to the frequency bands and pulse lengths used in microwave radar was practically nonexistent. Many types of commercially built test apparatus were pressed into service, and some of these were used continually through the war. However, completely new apparatus was often necessary. At first, the trend was toward production of many specialized test units, designed to perform specific test functions with one particular radar system. Cables and fittings were provided that permitted direct connection of the unit into the radar system. While in most cases this increased the ease of servicing the radar, it also led to the production of many different units having basically similar characteristics. Later, considerable effort was made to incorporate general-purpose functions in new test equipment designs and to make the units applicable to as many systems as possible.

17·2. Measurement Problems.—Measurement of physical quantities requires the availability of measurement standards. However, absolute instruments are seldom necessary or desirable outside calibration laboratories. In most electronic design and maintenance, secondary standards supply the basis for time measurements. The amplitude of an electrical quantity such as current, voltage, power, or flux density must usually be related to such secondary standards as precision meters, impedances, and standard cells. The bridge in its many forms is almost universally used for accurate impedance measurements, although direct use of frequency and amplitude standards may sometimes be desirable.

Comparison of an unknown quantity with the standard requires the use of some type of indicator. Of the various types used, the cathode-ray tube is by far the most versatile. The electron beam requires very little power for deflection because of its low inertia. The orthogonal character of the usual deflection system allows the introduction of at least two variables, and intensity modulation of the beam allows the use of a third if desired. Other visual and aural devices that will respond to an electrical signal find application as indication means and are used where particularly adapted.

Probably the most general and important measurement function in electronic equipment is waveform analysis. Here, simultaneous measurement of both time and signal amplitude is required. Meters will indicate the peak, average, or effective amplitude; frequency-selective devices will give the harmonic content; but the oscilloscope or oscillograph will show the characteristics of the complete waveform as a function of time. At frequencies higher than say 5000 cps the cathode-ray tube becomes the only generally applicable indicator for this type of presentation.

Determination of the frequency-response characteristic for all types of electrical networks and amplifiers presents another type of measurement problem. Classic point-by-point methods using a c-w signal generator and indicating meter are probably the most accurate but are very time-consuming. Techniques employing f-m signal generators permit direct presentation of the entire characteristic on a cathode-ray tube screen and make adjustment of circuit components much quicker and easier.

Electronic circuit design must be based on a broad knowledge of the characteristics of electrical circuit components. These must be determined accurately for the operating conditions under which the component is to be used. A wide variety of accurate and reliable commercial instruments is available for making impedance measurements of simple linear elements and determining characteristic curves for nonlinear elements under d-c conditions. Component characteristics may be considerably different, although operated within their power ratings, when a high-amplitude, low-duty-cycle pulse waveform is applied. Skin effects and capacitive and inductive effects become important in the presence of high-frequency components in the pulse waveform. Operation of vacuum tubes in highly nonlinear regions of their characteristics also complicates both measurement and circuit design problems. Some special equipments have been designed to measure component characteristics under pulse conditions; but as these conditions vary widely, simple pulse voltage and current measurements or substitution measurements are usually employed.

17·3. Laboratory Equipment.—Almost all types of commercially available electrical test equipment were used in the radar development laboratories. Such units as signal generators, vacuum-tube voltmeters, multimeters, bridges, decades, wave analyzers, and "Q" meters were required in large quantities and used with very little modification. Commercial oscilloscopes were used directly for servomechanism and other low-frequency design work and, with the addition of an external synchronizer unit, for the design of many pulse circuits.

Much equipment of entirely new design was necessary for the development of radar receivers, indicators, and ranging equipment. Included in this group were new types of cathode-ray oscilloscopes for indicators, apparatus for accurate time measurement, pulse generators for simulating signals and making impedance measurements, and signal generators for determining the bandpass characteristics of circuits. These equipments underwent parallel development with the radar systems. New radar techniques and circuits were incorporated in test equipment design as developed.

The mechanical requirements of test equipment designed strictly for laboratory use are not particularly stringent. It is not subjected to heavy shock, vibration, or large variations of temperature and seldom has to be moved appreciable distances. In addition it is usually operated and maintained by skilled technical personnel. Consequently, other factors can be subordinated to the attainment of high accuracy, ease or rapidity of construction, and convenience of use.

17·4. Field Equipment.—The complexity of radar apparatus soon made adequate field test equipment an absolute necessity. Maintenance crews found that "a neon bulb and a screwdriver" were not sufficient to keep a radar system operating at all, much less at maximum efficiency. The necessity of providing test equipment for field use under adverse conditions presented a huge problem in design and production engineering.

All of the previously mentioned types of laboratory equipment were necessary for use in the field along with many additional units to meet the special requirements of the Services. These units were often required to be as accurate as the equivalent laboratory equipment even when operated under field conditions by nonscientific personnel. Field conditions include the extremes of temperature, humidity, shock, and vibration.

Satisfactory operation at the extremes of ambient temperature ($-55°$ to $+71°C$ in some specifications) required development of new components and types of insulation, the use of forced air cooling, and operation of components at much less than their room-temperature voltage and power ratings. To meet humidity and immersion specifications, hermetic sealing of all critical components and the use of corrosion-resistant metal finishes were required. In some cases, entire equipments were built in waterproof containers. Completed chassis were sprayed with special varnish to resist the growth of tropical fungus. Special attention was required on bakelite, wood, fiber, and leather in this regard.

Careful consideration of the effects of shock and vibration is required in the mechanical design of field test equipment. Since most of these equipments must be portable, the use of shock-absorbing mountings is usually impractical. Transportation over rough roads in vehicles having stiff springs is the rule rather than the exception. The added weight of sealed components further increases the mechanical design problem. Chassis must be braced and reinforced, and components securely clamped to meet these requirements. At the same time, weight restrictions are severe. It is impossible to meet the requirements imposed by the nature of field test equipment without keeping them in mind from the earliest stages of both mechanical and electrical design.

CHARACTERISTICS OF RADAR TEST EQUIPMENT

17·5. The Cathode-ray Oscilloscope.—The cathode-ray oscilloscope is an important direct-measurement instrument and the most versatile indicator yet developed for comparing electrical quantities. Probably its greatest single advantage is its ability to indicate instantaneous values of rapidly changing voltages. Electromechanical oscilloscopes have many of the advantages of the cathode-ray oscilloscope, but their inertia limits their frequency response to the a-f range.

The heart of the cathode-ray oscilloscope is, of course, the cathode-ray tube. In this device, a beam of highly accelerated electrons passes through a deflecting field (electrostatic or magnetic) before striking a fluorescent screen at the end of the tube. The deflecting field is usually the resultant of two mutually perpendicular fields which can be independently controlled by signal voltages. If one of these voltages increases linearly with time, the waveform (time variation) of the other voltage will be graphically indicated. Since most electrical quantities can be converted into voltages having proportional variation, they may in this way be directly or indirectly observed.

Knowing the deflection sensitivity of the cathode-ray tube, voltage measurements may be made with a scale on the screen. Accuracy is limited by the manufacturing tolerances on the cathode-ray tube, which limit the linearity of deflection, and by the difficulty of reading the dot or trace position accurately. The diameter of the dot when focused on the screen may be about 0.03 in. at low intensity, increasing in size as the intensity of the electron beam is increased. These factors limit the possible accuracy of direct measurement to about 1 per cent.

Considerably greater accuracy may be obtained when the cathode-ray oscilloscope is used as a null indicator for comparing the quantity to be measured with a standard. It may, in fact, be limited only by the accuracy of the standard used. Measurements of time, frequency, phase, harmonic distortion, impedance, and other related quantities may be made, using appropriate auxiliary apparatus.

17·6. Cathode-ray Tube Development.—The development of the cathode-ray tube was greatly accelerated during the war, and many new and improved designs of magnetic and electrostatic deflection tubes were produced. Magnetic deflection tubes have found application mainly in radar plan-position indicators (PPI's), while electrostatic deflection types are used for radar "A-scopes" and in test equipment.

New electrostatic deflection tubes were developed to meet the requirements of pulse circuit work for better definition, higher intensity, and lower input capacitance to the deflecting plates. Of the cathode-ray tubes most used in test equipment, the type 902 with a 2-in. screen was replaced by the 2AP1, and the 3AP1 with a 3-in. screen was replaced by the 3EP1/1806 and later by the 3BP1 with a medium shell diheptal base. The 5BP1 has been largely superseded by the 5-in. 5CP1 having a medium shell diheptal base and a third accelerating anode which may be operated at twice the acceleration potential of the earlier tube. The high-deflection-sensitivity type 5LP1 will probably be replaced by the 5JP1 which has much lower input capacitance to the deflecting plates. In addition, 3- and 5-in. radial deflection tubes were developed for use with circular sweeps. The type 3DP1, having an electrode in the center of the screen, is now in general use.

The P1 screen is the most desirable type now available for direct viewing. The high sensitivity of the human eye to a green trace permits operation at lower beam current than is possible with white, blue, or amber trace screens. This, in turn, permits better focus and definition. The medium persistence time is also well suited to the average measurement problem.

Blue trace P5 and P11 screens are well suited for direct photography of the tube face. Their light spectra coincide closely with the region of maximum sensitivity of ordinary photographic film. The P5 has a somewhat shorter persistence time than the P11 and may be used for continuous film recording of waveforms at frequencies up to 60 kc/sec.[1]

Waveforms having frequencies lower than 20 cps make a long-persistence screen very desirable. The P7 screen, developed for radar indicators using magnetic deflection tubes, has also been supplied in standard electrostatic deflection cathode-ray tubes. Such a tube is very useful for viewing transients which may remain visible for as long as 2 min.

For best definition with any of the above-mentioned screens, color filters and hoods are desirable. A plastic color filter alone will make viewing possible under bright ambient light conditions. When mounted within a hood at an angle of 30° to 45° with the tube face, reflections from both the tube and the filter will be eliminated. A light-tight hood permits operation of the cathode-ray tube at very low intensities and gives the best possible definition. However, lighted grids or scales are necessary if direct voltage measurements are to be made. Edge-lighted plastic plates with ruled grids have been found to be very useful, for by dimming the light, the effect of the superimposed grid can be almost eliminated.

The basic cathode-ray oscilloscope consists only of a cathode-ray tube and a high-voltage power supply containing intensity, focusing, and centering control. The majority of the newer types of measurement problems to which the cathode-ray oscilloscope is applicable requires the use of supplementary electronic apparatus. This is particularly true where time measurements are necessary.

17·7. Deflection Systems.—Viewing an electrical waveform without distortion requires that the horizontal and vertical deflection of the elec-

[1] H. Goldstein and P. D. Bales, "High Speed Photography of the Cathode Ray Tube," *Rev. Sci. Inst.*, **17**, No. 3, 89–96, March 1946.

tron beam be directly proportional to the amplitude and the independent variable of the waveform. Since the independent variable is usually time, some type of "sawtooth" waveform is used to sweep the beam across the tube face at a constant rate periodically. Other waveforms, including sinusoidal, parabolic, and hyperbolic, are sometimes required for special applications.

The use of a linear-sweep voltage makes the cathode-ray oscilloscope applicable to the field of measurement in which a graphic presentation of the measured quantity is necessary. The most common application is signal tracing in electronic equipment in which a qualitative knowledge of the waveforms present is adequate. Quantitative measurements of waveform amplitude as a function of time may be made to an accuracy of about 1 per cent by direct measurement on the cathode-ray tube face. A much higher degree of accuracy may be obtained using the linear-sweep presentation as a null indicator for comparison of the signal with time and amplitude standards.

The circular sweep has more limited but very important applications. It is particularly desirable for frequency comparisons and precise time measurement. The fact that no retrace is necessary and that a much longer baseline is obtained (6 to 8 in. on a 3-in. tube) tends to offset the disadvantage of waveform distortion with polar presentation.

Frequency-comparison measurements may also be made by applying a sinusoidal sweep voltage directly to the horizontal deflecting plates. The equal-frequency Lissajous figure is readily recognizable for both c-w and pulse waveforms, although with the latter the adjustment problem is fairly difficult.

Most other types of sweeps have little application in test equipment. Intensity-modulated television sweeps may be used for cathode-ray tube testing or for specialized television applications. Other special applications may require hyperbolic or parabolic sweep voltages. Spiral time bases offer some additional possibilities. Treatment of these and other types of presentation will be found in Vols. 19 and 22 of this series.

Two forms of linear-sweep generators are in general use. One is a conventional free-running sawtooth generator of variable frequency. Synchronism between signal and sweep is obtained by first making the sweep frequency slightly lower than the signal frequency (or a submultiple of the signal frequency). A portion of the signal voltage is then introduced into the sweep generator, causing it to fire slightly sooner on each cycle and locking it in phase and frequency with the signal. This type of circuit is very useful for viewing continuous waveforms but requires constant attention to the synchronism and sweep frequency controls when a series of different waveforms are to be viewed. Synchronization with low-duty cycle pulse waveforms may be quite difficult or impossible.

A second sweep generator is the start-stop type which produces a single excursion of the electron beam for each synchronizing signal applied. When synchronizing trigger pulses are used to start both the sweep and the waveform to be viewed, no synchronizing adjustment is necessary. Since most radar waveforms are initiated by triggers to make precise timing possible, this type of sweep generator has found general application. Most viewing requirements can thus be met by a number of fixed sweep durations selected by a tap switch.

Both types of sweep generators consist of integrating devices in which a condenser is charged or discharged by an essentially constant current. Usually an electronic switch tube is used to discharge a condenser which is then permitted to charge again from a constant-current source. The free-running sweep generator is usually a relaxation oscillator which rapidly discharges the condenser when the voltage across it exceeds a predetermined value. The triggered, single-stroke type of sweep generator employs a switch tube having a stable, quiescent condition. Upon introduction of a trigger pulse, the switch tube is cut off ("opened") for a definite period of time, permitting the voltage across the condenser to rise. Conventional methods of obtaining constant-current charging, low-impedance output, and push-pull voltages are applicable to both types of circuits.

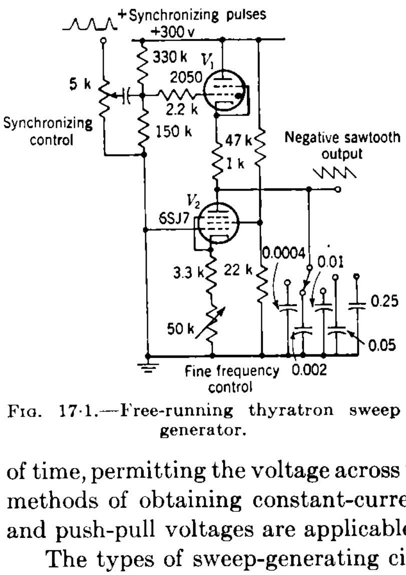


FIG. 17·1.—Free-running thyratron sweep generator.

The types of sweep-generating circuits that have been developed are described fully in Vol. 19, Chap. 10, in Vol. 22 of this series, and in *Time Bases* by Puckle.[1] Examples of the practical application of most of these circuits will be found in the following chapter. It will suffice to mention here the various types of circuits applicable to test equipment.

The thyratron is the most widely used free-running switch tube. However, the maximum sweep frequency obtainable is limited to between 40,000 and 50,000 cps by the time required for deionization to occur. A typical circuit is shown in Fig. 17·1. Vacuum tubes may be used for free-running sweep generators at frequencies up to 1 Mc/sec and for triggered switching applications. A blocking oscillator is used as a free-running switch in Fig. 17·2, while Fig. 17·3 includes the gated triode switch most commonly used in test equipment.

[1] O. S. Puckle, *Time Bases*, Chapman & Hall, Ltd., London.

A number of methods of producing a linear rather than an exponential rise of voltage have been developed. The simplest of these uses a very

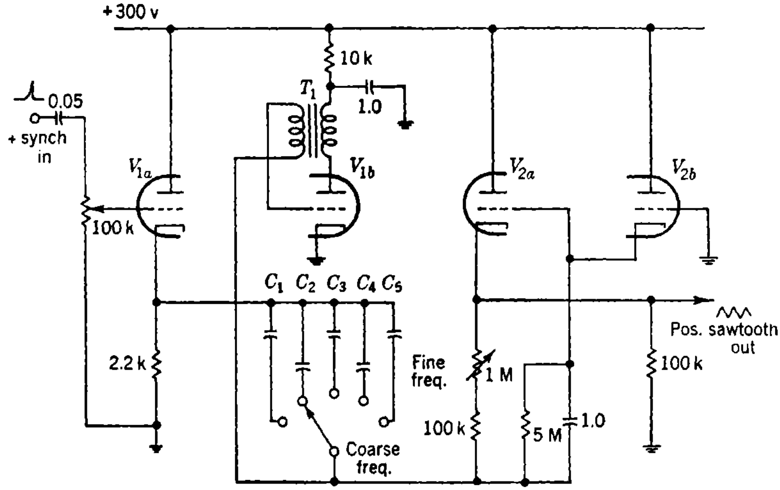


FIG. 17·2.—Free-running blocking-oscillator sweep generator.

small portion (5 per cent or less) of the characteristic exponential condenser charging curve. An effective *RC* time constant roughly twenty times the expected sweep duration will ensure this. Another method is to use a pentode as the charging resistance, taking advantage of its very high plate resistance to provide constant-current charging as shown in Fig. 17·1. These methods will be found generally in oscilloscopes employing a free-running thyratron switch tube.

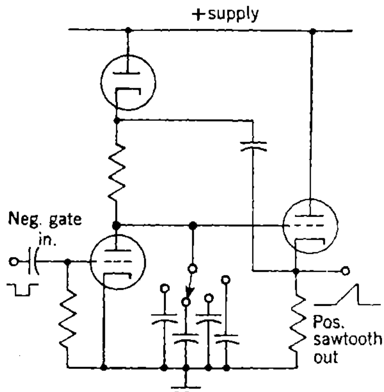


FIG. 17·3.—Sweep generator using a triode switch tube and positive feedback for linearization.

A positive-feedback unity-gain amplifier may be used to maintain a constant voltage across the charging resistor as the condenser voltage rises. The simplest form of this circuit uses a cathode follower as shown in Fig. 17·3. This type of circuit is widely used in radar test equipment. A negative-feedback amplifier may also be used to make a very stable linear-sweep circuit. A circuit of this type is shown diagrammatically in Fig. 17·4.

Approximate linearization may be obtained by inserting a large inductance in series with the charging resistance as shown in Fig. 17·5. This method reduces the required tube complement but introduces severe duty-cycle limitations. It will be found in a number of the oscilloscopes described in Chap. 18 but has been largely replaced in new designs by one of the feedback types of sweep voltage generators.

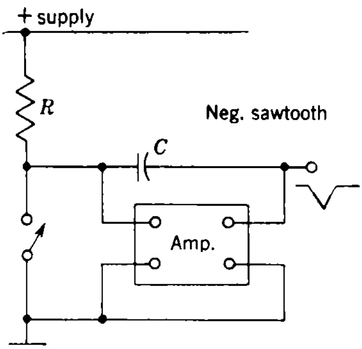


FIG. 17·4.—Sweep generator using negative feedback for linearization.

The voltage swing required of any one tube may be halved and uniformity of focus materially improved if the beam deflection is accomplished by push-pull voltages applied to opposite plates. An inverter amplifier is required, which may be a single-stage negative-feedback amplifier as shown in Fig. 17·6. When lower maximum voltages are required, a push-pull amplifier of the cathode-coupled type shown in Fig. 17·7 may be desirable. If the series inductance type of sweep generator is used, push-pull voltages having about 80 per cent balance may be obtained using a center-tap choke as shown in Fig. 17·5.

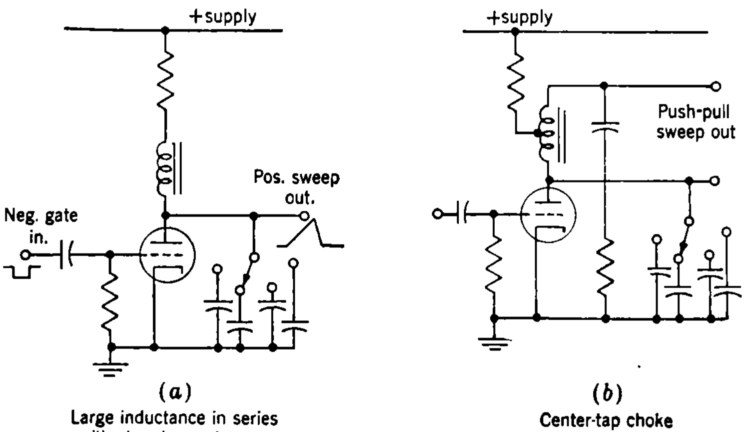


FIG. 17·5.—Sweep generators using a high-impedance choke for linearization.

A circular sweep is generated when sinusoidal voltages that differ by 90° in time phase are introduced onto deflecting plates that have a 90° space-phase relationship. Probably the simplest method of obtaining these voltages is to use the voltage drops produced across a resistor and condenser when connected in series across an alternating current source.

If good focus is to be obtained, push-pull voltages must be supplied. These are more accurately obtained by using a resonant transformer as shown in Fig. 17·8. When the secondaries are tuned on opposite sides

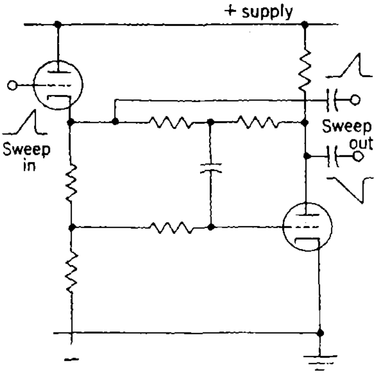


FIG. 17·6.—Negative-feedback inverter.

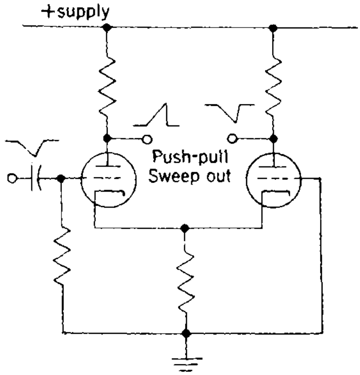


FIG. 17·7.—Cathode-coupled inverter.

of resonance, both phase and amplitude differences are obtained. By simultaneous adjustment of the secondary tuning capacities C_2 and C_3, equal deflections in both coordinates and a 90° phase difference between secondary voltages may be obtained. Circle diameter may be controlled by adjusting the signal amplitude or by tuning off resonance with the primary tuning capacity C_1. Use of high-Q components aids in eliminating distortion from the sweep voltages and produces a more nearly linear sweep rotation rate. By inserting this tuned circuit in the plate circuit of a crystal oscillator or buffer stage a very accurate time base may be generated quite simply.

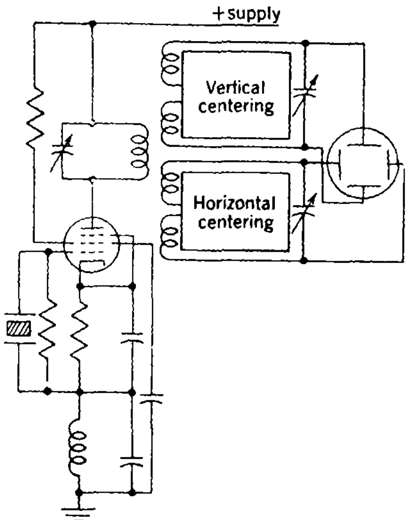


FIG. 17·8.—Pentode crystal oscillator and phase-shift transformer for generating a circular sweep.

17·8. Signal Channels.—The cathode-ray tube has one property that is particularly valuable in a general-purpose measurement instrument, that of high input impedance. The capacity from any deflecting plate to the other electrodes is in almost all cases less than 10 $\mu\mu$f so that the input impedance at all frequencies up to 15 Mc/sec is greater than 1000 ohms. For a low-capacity tube such as the 5JP1 (which has an average of 3.8$\mu\mu$f from a deflecting plate to all other elements) the direct input impedance is found to be greater than 1000 ohms up to 40 Mc/sec. Even lower input capacitances are obtained with the type 3DP1 tubes, a typical value being 1.8$\mu\mu$f from center electrode to all other electrodes.

A major disadvantage of the cathode-ray tube is its low voltage sensitivity. This presents no particular problem where signals of greater than 5 volts amplitude are to be displayed, as is the usual case in radar synchronizing, indicating, and ranging equipments. Use with small signals such as are found in some radar signal circuits requires amplifiers that will not distort the waveform. These amplifiers must have small amplitude and phase distortion in the band between the lowest- and highest-frequency components of the signal if a true picture of the waveform is to be obtained.

The design of wide-band "video" frequency amplifiers is one of the most critical steps in the design of test equipment. Very careful consideration of compensation for the effects of loading due to tube and wiring capacities and of mechanical layout of components is necessary to obtain optimum gain and bandwidth with a given power input. For cathode-ray tube deflection this is of particular importance, since an undistorted output of 30 to 100 volts is normally required.

A number of requirements must be considered. The rise time of the amplifier when a step function is applied is of more direct interest for pulse applications than the bandwidth. It is, in general, a function of the bandwidth and may be approximated by the expression

$$\text{Rise time} = \frac{0.35}{f_h}, \tag{1}$$

where f_h is the frequency at which the mid-frequency gain is down by 3 db. The rise time is usually measured as the time required for the pulse to rise from 10 to 90 per cent of its final amplitude, since these points are fairly well defined.

Overshoot in the output pulse represents a form of distortion. It is a measure of the amount that the output signal initially exceeds its nominal amplitude and results from a nonlinearity in the phase-shift vs. frequency characteristic of the amplifier. The amount of overshoot is normally limited to less than 5 per cent in amplifier design.

The ability of the amplifier to sustain a long rectangular waveform is related to the low-frequency gain characteristic. Normally, about a 10 per cent drop in the pulse amplitude is the maximum permitted in the longest pulse that the amplifier is designed to pass. The amount of slope in the top of the output waveform is directly dependent on the time constants in the cathode, screen grid, and plate circuits and in the grid and plate coupling circuits. These time constants are made as large as is practical without impairing the high-frequency response by capacity-to-ground of large components or unduly increasing the recovery time. Actual selection of values usually requires a rather careful compromise between these conflicting requirements.

The signal delay introduced by the amplifier must be considered in timing applications. The delay will be the same for all frequency components of the signal in the region where the phase-shift vs. frequency characteristic is linear. This region corresponds closely to the region of "flat" frequency response.

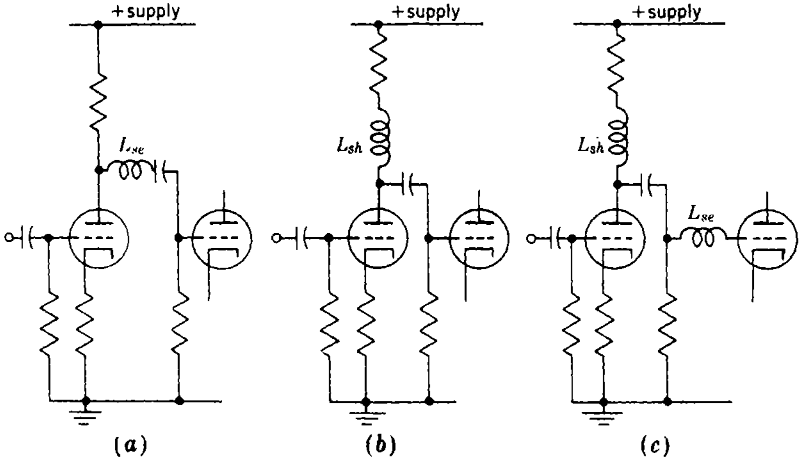


FIG. 17·9.—Basic peaking circuits. (a) Series peaking; (b) shunt peaking, (c) series shunt peaking.

A treatment of the design methods used to meet the above requirements is beyond the scope of this chapter, but will be found in Vols. 18 and 23 of this series.

The tubes used in these amplifiers are chosen by a figure of merit which is the ratio of the transconductance to the total of the tube capacitances. For a given tube, with its associated socket and wiring capacities, the plate-load resistance is then determined by the bandwidth requirements:

$$f_h = \frac{1}{2\pi R_L C_T}, \tag{2}$$

where C_T is the total of all shunt capacities in the plate circuit. With the load resistance and g_m set, the mid-frequency gain is determined. The tube choice must have been made such that the rated power dissipation of the tube will not be exceeded when the operating point is set for the required output voltage swing. The stage gain is also set by the determination of R_L and g_m.

The bandwidth of the amplifier may be increased somewhat by the use of series and/or shunt peaking. The basic peaking circuits, shown in Fig. 17·9, consist of an inductance in shunt or in series with the capacity to ground or a combination of the two. In any case, the small inductance included is adjusted to resonate with the shunt capacity at a somewhat higher frequency than f_h without peaking. This effectively extends the region of flat frequency response by a factor between 1 and 1.7. Over-peaking, setting the resonant frequency low enough to cause an increase in stage gain at high frequencies, is to be avoided, as it results in over-shooting. Other considerations such as plate circuit decoupling to avoid feedback are essentially the same as required for conventional high-gain a-f amplifier design.

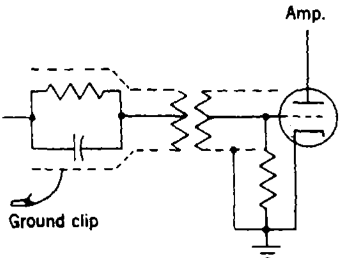


FIG. 17·10.—Voltage divider probe.

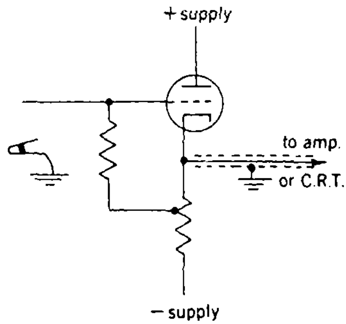


FIG. 17·11.—Cathode-follower probe.

Since a few micromicrofarads of amplifier input capacitance may represent a very considerable loading of the circuit under test, signal probes are often used which reduce this capacity to a minimum. Two types are in use. In one type, a resistance-capacitance divider is inserted at the end contact of the shielded probe lead which makes the connection from the circuit under test to the amplifier. This divider, shown in Fig. 17·10, reduces the capacity and attenuates the signal by an approximately equal factor so that a higher-gain amplifier is required than would be necessary otherwise. The constants are adjusted so that $R_1C_1 = R_2C_2$, where C_2 represents the capacity of the lead plus the input capacity of the amplifier. The reduction in capacity by this method is limited only by the point at which the bandwidth reduction introduced by the additional amplifier stages exceeds that produced by the capacity of the probe.

The other method of reducing input capacity is to mount the first stage of the video amplifier in the probe itself. This stage is designed to have an output impedance low enough that the effect of the capacity of the lead and the input capacity of the following stage is negligible. A cathode follower is desirable for this application because of its low output impedance. Furthermore, if the gain of the cathode follower is G, the input impedance due to the grid circuit can be increased by a factor approaching $G/(1 - G)$ by returning the grid resistor to the cathode as shown in Fig. 17·11. A circuit described in Sec. 18·11 uses another cathode follower to make the plate supply vary in the same manner as the cathode, keeping the "Miller" grid capacity to a minimum.

The resistance-capacitance divider is also often used for gain control. Accurate gain calibration can be obtained by selecting one of a series of these dividers with a tap switch. A calibrated attenuator may also be used to indicate a difference in voltage and power levels. This is obtained directly from the difference in attenuator settings for equal signal amplitudes at the output of the amplifier.

17·9. Auxiliary Circuits.—The usefulness of the cathode-ray oscilloscope may be increased by including measurement standards and other auxiliary circuits in its design. These include time and voltage standards, time and voltage comparison circuits, synchronizing pulse generators, circuits for mixing two or more signals on a single trace, and circuits for presenting signals simultaneously on two or three traces.

Either pulsed or continuous oscillators may be included as time and frequency standards. These may be LC or crystal controlled, depending on the accuracy required. The output may be presented directly as sinusoids or converted into marker pulses. Calibrated delay circuits may be used as time standards or as interpolating devices. The usual voltage standard is a meter reading directly the d-c voltage required to deflect the baseline on the cathode-ray tube a distance equal to the amplitude of the signal being measured. A precision potentiometer and voltage difference amplifier may be used for more accurate measurements.

Since many types of circuits require externally generated synchronizing pulses to initiate their action, a trigger generator is often desirable in the oscilloscope. When continuous oscillators are used for timing, the trigger pulse is normally locked in phase with the oscillator voltage by some form of frequency division. This permits simultaneous synchronization of the oscillator with the signal and the sweep.

When two signals are to be compared in this manner on a single time base, methods of mixing them must be provided. The simplest method of doing this is to connect the signals to opposite deflecting plates in the cathode-ray tube. The resultant coupling of the two circuits is usually negligible because of the low capacity between plates. Where more signals must be superimposed or push-pull deflection is necessary, tube or switch mixers must be used. Signals may be applied to both the grid and cathode of a triode amplifier tube with the mixed output appearing at the plate. Pentode and pentagrid mixer tubes will combine more signals in a single stage. Additional signals may be applied to subsequent stages, or a common plate or common cathode-resistor may be used with a number of tubes, giving an output proportional to the sum of the signals applied to the grids. Representative circuits are shown in Fig. 17·12. Some of the voltage adding circuits discussed in Chap. 3 and in Vol. 19, Chap. 18, may also be used for combining a number of signals.

When the distortion caused by superimposing two or more signals is undesirable, electrical or mechanical means of switching the deflection

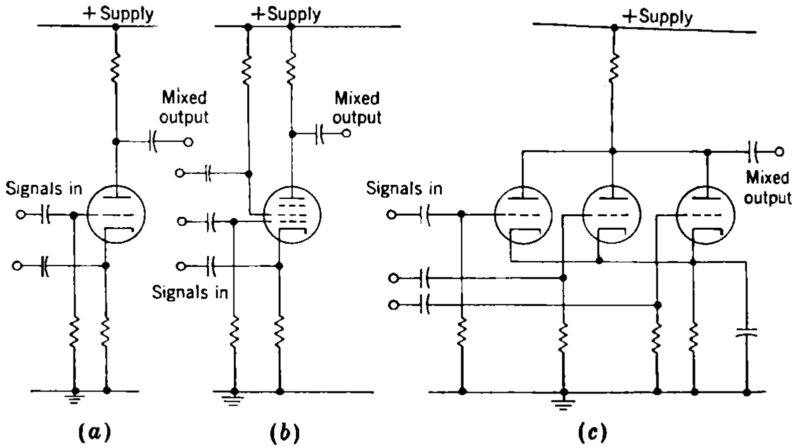


FIG. 17·12.—Mixer circuits. (a) Triode mixer; (b) pentode mixer (screen input should have low impedance source); (c) common plate mixer.

plates to each of the signals on consecutive sweeps may be used. A motor-driven multiple switch or capacitive coupler is entirely practical if it can also supply synchronizing pulses for the signal and sweep circuits. Electronic switching circuits of the multivibrator or "flip-flop" (scale-of-two) type are also used. A circuit of this type for use with triggered sweeps is shown in Fig. 17·13.

Another approach to this problem uses a cathode-ray tube having multiple gun and deflecting plate structures. Tubes of this type are available from several manufacturers, although as this is written, only the two-gun type 5SP1 has received RMA nomenclature. Two- and three-gun tubes in both 3- and 5-in. screen diameters have been designed. The sets of deflecting plate for each gun structure are shielded and are brought out to separate connections. Identical sweep voltages may be used when waveform comparison is desired, or different sweep voltages may be applied for two-scale timing systems or similar applications.

17·10. Timing Oscillators.—Oscillators in any of a large number of forms have considerable application as pieces of test equipment. They may be used as standards of time or frequency or as signal sources. They may also determine repetition rates or the rate of any other recurrent phenomena. They may have their frequency determined electrically by an RC- or LC-circuit or mechanically by the characteristics of a quartz crystal, vibrating fork, or motor. Their output may be sinusoidal,

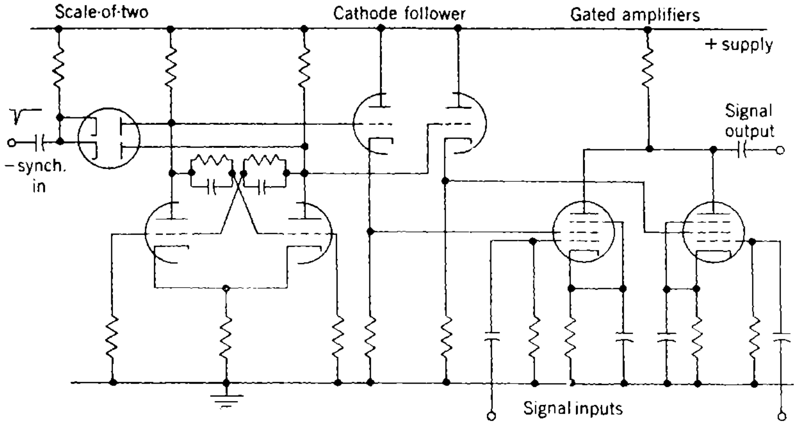


FIG. 17·13.—Electronic switch for video signals.

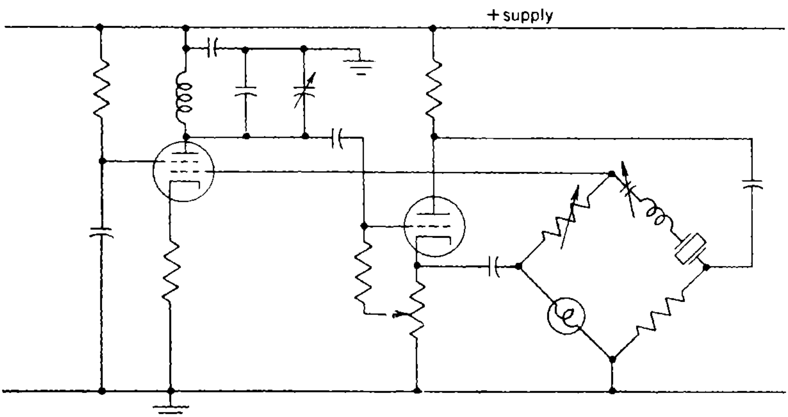


FIG. 17·14.—Bridge stabilized crystal oscillator. (*Courtesy of General Radio.*)

square, triangular, or a train of pulses. They may oscillate continuously or for a given length of time after receiving a synchronizing pulse.

The crystal oscillator has the highest inherent stability and is generally used as a secondary frequency standard. For any given circuit the crystal may be ground to almost the exact frequency of a primary standard and then "pulled" to the exact frequency. A crystal will drift only

1 to 3 cycles per megacycle per degree if cut to obtain a low-temperature coefficient. The oscillator circuit used should be designed to reduce the frequency shift produced by tuning the plate circuit and to give a constant output voltage. A representative circuit meeting these requirements is shown in Fig. 17·14 and is described in the April 1944 issue of the *General Radio Experimenter*. Where an accuracy of about 0.1 per cent is adequate, a simple triode oscillator with no regulation or compensation will suffice.

Oscillators using *LC*- or *RC*-circuits as their frequency-determining elements may be easily designed to give a variable output frequency. Single-frequency accuracy comparable with that of a crystal oscillator may be obtained by very careful design.

The oscillator circuits employed for the generation of timing pulses have been described rather completely in Vol. 19, Chap. 4, and examples of their application will be found in Chap. 18 of the present volume. Some discussion of the frequencies used and types of output signals generated is appropriate for this section.

In radar applications it has been convenient to measure time directly in terms of an equivalent radar distance. On this basis 1 μsec is approximately equivalent to 164 yd. Oscillator frequencies in general use range from 1617 cps for a 50-nautical-mile period to 1.639 Mc/sec for a 100-yd period. Oscillator frequencies in use in test equipment include 80.86 kc/sec (1 nautical mile), 81.94 kc/sec (2000 yd), 93.11 kc/sec (1 land mile), 100 kc/sec (10 μsec), 166.88 kc/sec (1000 yd), and 819.4 kc/sec (200 yd). Multiples or submultiples of these frequencies are obtained by frequency multiplication or division. For general-purpose instruments, however, standardization on frequencies giving direct readings in microseconds is desirable. Most of the 80- or 90-kc/sec oscillators in existing equipments may be tuned to use 100 kc/sec crystal by a small reduction in the tuning capacity. A similar change can be made for most of the other frequencies that may be encountered.

Oscillators generating a sinusoidal waveform are usually used in timing applications. A large number of cycles must lie in the interval of time being measured if accuracy is to be obtained. This introduces counting difficulties and may cause a superimposed signal waveform to appear distorted. Squaring the sinusoidal wave by amplification and clipping gives a fast wavefront which permits more accurate reading. Peaking this wave with an *RC*-circuit having a short time constant gives a train of pips that cause less distortion of the waveform on which they are superimposed. A blocking oscillator may be synchronized by pips generated across a small choke, damped by a resistance, in the oscillator plate or cathode circuit. This will generate the desired waveform directly.

When higher accuracy is required than can be obtained by visual interpolation between a reasonable number (10 to 30) of marker pips, interpolating circuits must be included. One method is to use a calibrated phase shifter to vary the phase of the pips with respect to the synchronizing trigger. Another is to provide a movable marker pulse or a second scale of markers. These are used with an expanded oscilloscope sweep to obtain maximum reading accuracy.

17·11. Delayed-pulse Generators.—The delayed-pulse generator may be used for time and distance measurements and in setting up timing sequences. This circuit accepts a pulse and produces another pulse after a period of time determined by the circuit constants. The delay time may be fixed or variable, depending on the requirements of the application. When variable, the control may be calibrated to read directly in delay time.

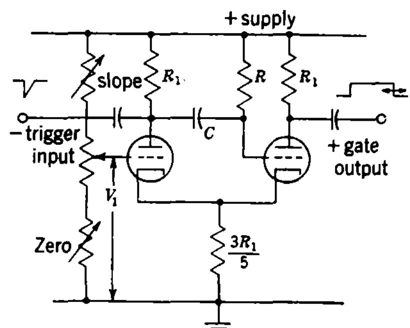


FIG. 17·15.—Delay multivibrator.

Stability of delay time from cycle to cycle and over a period of time is important. Variations in the delay time of succeeding pulses introduces a time variation or "jitter" in all of the following circuits or in the delayed pulse on an oscilloscope screen. This may upset the operation of these circuits and will reduce the accuracy of time measurements by producing a broad trace. This jitter cannot be completely eliminated but can be reduced to a negligible point for a given application by careful circuit design.

When a calibrated variable delay is desired, the ease of calibration must also be considered. A linear relation between the rotation of the control shaft and the time delay produced facilitates this process and makes special dials unnecessary. Slope and zero controls must be adjusted to make the indicated and actual delay "track." It is desirable to have the action of these controls independent and to eliminate either where possible.

Of the large number of delay circuits developed for radar applications only a few have found application in test equipment. Those commonly used are the delay multivibrator, the phantastron, and the linear-sweep delay. Blocking oscillator pulses are sometimes used for short fixed time delays as are lumped constant delay lines.

The delay multivibrator is the least stable of these but requires the smallest tube complement. A representative circuit, shown in Fig. 17·15, requires only two triode tube sections. The delay time is largely determined by the time constant RC, with the circuit values normally used, and is sensitive to variation in pulse repetition rate, plate and heater voltages, and tube changes. This delay circuit if frequently calibrated may be used for interpolation with an accuracy of better than 1 per cent, when a regulated power supply and constant repetition rate are used. Its delay may be made linear with variation of the voltage V_1 to about 0.2 per cent.

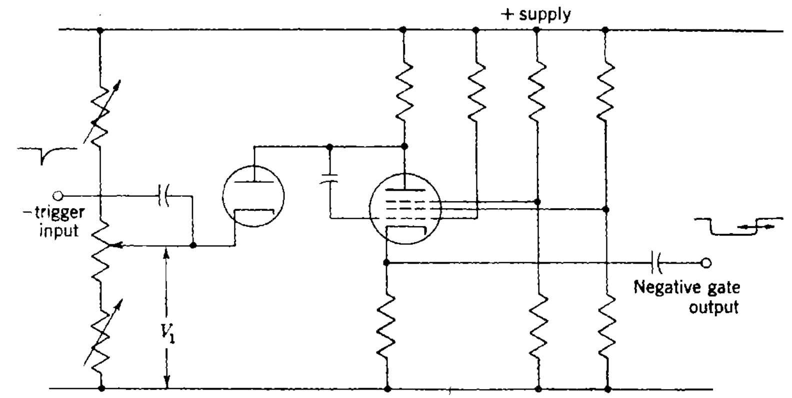


FIG. 17·16.—Phantastron delay.

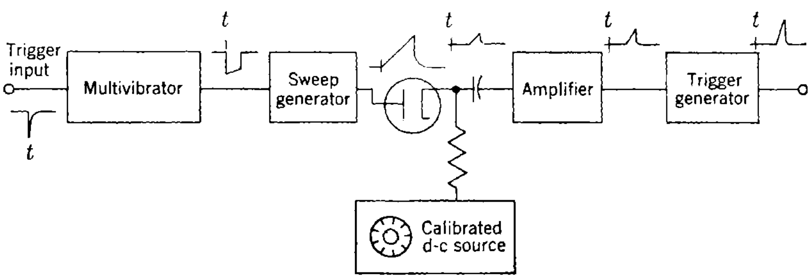


FIG. 17·17.—Linear-sweep delay block diagram.

The phantastron and linear-sweep delays may be made considerably more stable than the delay multivibrator but require more tubes. The phantastron, for which a circuit is shown in Fig. 17·16, requires a pentode with sharp cutoff suppressor characteristics (a 6AS6 or 6SA7 is used), a diode, and one or more amplifier stages. With careful design it may be made linear to 0.1 per cent with variation of the voltage V_1 and, with occasional calibration, accurate to about 0.3 per cent. Any of the linear-sweep generating circuits mentioned in Sec. 17·7 may be used with a "pick-off" diode (through which a control potential is introduced) and one or more amplifier stages. An average of six to eight tube sections is required. A block diagram of this type of circuit is shown in Fig. 17·17. These and other time-modulation circuits are described in detail in Vol. 20.

Fixed and variable time delays of from 1 to 10,000 μsec duration are in general use in radar systems and test equipment. Many of these circuits may be used outside this range with minor redesign.

17·12. Video Pulse Generators.—Both frequency- and pulse-response of the video amplifier are of interest. An approximate relationship between the bandwidth and pulse rise time has been pointed out in Sec. 17·8. The frequency-response characteristics do not immediately give the complete transient response of the circuit to a pulse, however, since the phase-shift characteristic is not directly indicated and the latter cannot be computed unless a minimum phase-shift network can be assumed. The use of rectangular pulses directly for video-amplifier testing gives much more information on the pulse amplification characteristics. High-frequency response may be determined from rise time measurements. The ability of the amplifier to sustain a long pulse is also directly measurable and is a function of the low-frequency response of the amplifier. Any transient distortion is directly indicated by this type of analysis. In addition, the nonlinearities that can be tolerated in pulse circuits make ordinary bandwidth measurements of less meaning than direct pulse waveform observations, for the representation of a time function by a frequency function is valid only for a linear system.

To give a significant indication, the output pulse from the pulse generator must be as nearly square as possible when connected to the stage under test. A rise time of 0.03 to 0.05 μsec for a ± 50-volt pulse is usually considered satisfactory for short pulses. Pulse lengths of 0.1 to 1500 μsec are normally required and are usually covered in two or more ranges by separate circuits. Gas tubes offer an easy way to obtain the fast switching action required for production of short pulses, although "hard" tubes may be used. A variety of commercial square-wave generators are available for testing response to pulse waveforms.

Stage gain under pulse conditions must also be measured. This is usually done by inserting a calibrated attenuator in the input to the amplifier. The difference in the attenuator readings for equal pulse amplitudes at the output of the attenuator and the output of the amplifier is a measure of the gain of the stage.

CHAPTER 18

PRACTICAL RADAR TEST OSCILLOSCOPE DESIGNS[1]

By H. J. Reed, Jr., A. H. Fredrick, B. Chance, and

E. F. MacNichol, Jr.

This chapter summarizes particular instruments that incorporate in their design the circuits and principles treated in Chap. 17. It consolidates information on many of the radar pulse test oscilloscopes developed during the war period. Much of this information is available only in reports and manuals that may be difficult to obtain or might not otherwise come to the attention of the reader.

The electrical specifications of these instruments are intended mainly to facilitate the selection of equipment for given applications. These or similar instruments may be available for some time from the Army or Navy supply centers or their surplus sales or from the manufacturing organizations. An attempt has been made, however, to present sufficient information for the construction of units having similar electrical characteristics when this becomes necessary. The general construction techniques used in the original instruments are shown in the photographs included. If actual mechanical and wiring layouts are desired, these must usually be obtained from the references cited. Circuit diagrams giving component values are included. The discussion of circuit operation is intended to aid one in understanding the relation among the various circuits used and also their adjustment. Particular attention is given to the functions of the various controls and adjustments. The detailed analysis of the operation and waveforms produced in the various basic circuits is necessarily left to other chapters and Vols. 19 and 20 of this series. For further data and specifications, references are given to the designer, the manufacturer, and the procuring agency.

Units included are representative of equipments that have been built during the war. Most of these circuits represent working designs that have withstood the test of at least limited production and use. It should be understood that in many circuits the particular stray capacities present have been important in determining some of the circuit constants given. These may not be duplicated with changes in layout and wiring some so that adjustment of component values may be necessary. Many

[1] All of Chap. 18 is by H. J. Reed, Jr., except Secs. 18·4, 18·8, and 18·9.

components and tubes in particular will usually not be held to wartime standards, broad as they were, and should be checked with care.

18·1. The P4 Synchroscope.[1] *Function.*—The P4 synchroscope was designed for use as a general-purpose oscilloscope for laboratory study of short pulses. Its principal application has been in the study of magnetron or radar modulator pulses and of beacon transmitter characteristics. It has also found application in the design of receiving, timing,

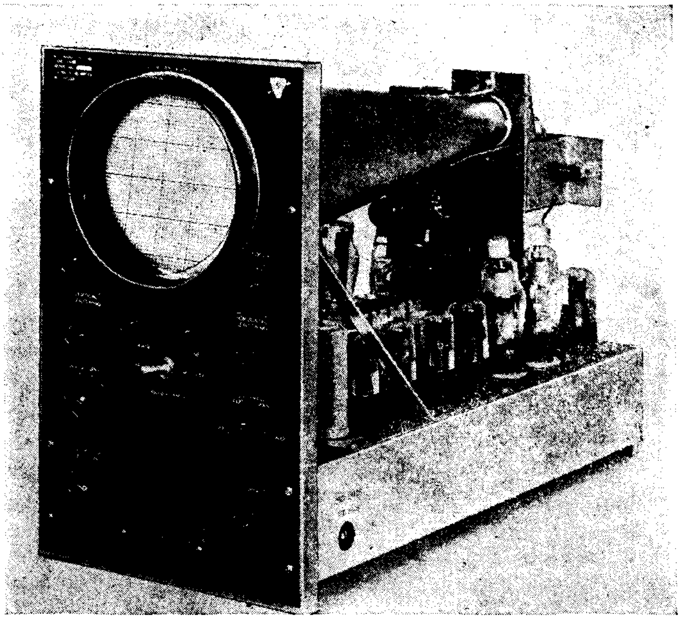


FIG. 18·1.—Synchroscope.

and indicating circuits, in the production testing of radar components, and as a general-purpose "A" scope for system testing.

Characteristics.—A type 5LPI cathode-ray tube operating at a total accelerating potential of 2600 volts is used, giving a vertical deflection sensitivity of about 60 volts per inch.

A triggered *sweep generator* of the constant-current pentode type is employed. Writing speeds are 0.04, 0.166, 0.5, and 2 in./μsec. Repetition rates provided by the local oscillator are 500, 1000, 2000, or 4000 cps. External triggering pulses having repetition rates between 200 and

[1] Developed and procured by the Radiation Laboratory. Manufactured by Sylvania Electric Products, Inc., and Browning Laboratories, Inc. RL Manuals M118 and M126 describe the complete equipment. The r-f detector and video amplifier are described in M124; the high-gain video amplifier is described in M166.

4000 cps or a sinusoidal oscillator in the same frequency range may be used. Either the internally generated trigger or an external signal of 10 to 300 volts amplitude may be used to synchronize the sweep generator. If an externally generated positive pulse of less than 100 volts amplitude is used, the intensification of the trace will be delayed about 0.2 μsec, and this portion of the observed pulse will not be seen. No sweep voltage outputs or inputs are provided.

The *trigger generator* output is a 135-volt positive trigger having a pulse width of 26 μsec between 10 per cent amplitude points. The pulse repetition frequency may be set by the local oscillator or to any frequency between 200 and 4000 cps by an external sinusoidal oscillator. Phasing of the output trigger with respect to the sweep trigger is variable by at least ±250 μsec. The cathode follower supplying the output trigger has an effective internal impedance of 500 ohms.

The only *signal channel* provided in the standard P4 synchroscope is a direct connection to the top vertical deflecting plate which has an input impedance of 1 megohm paralleled by 20 μμf or less, depending on the signal switch position used. Several auxiliary amplifiers have been designed to mount in this synchroscope. One type includes an r-f input fitting and an r-f square law detector which is used in examining the envelope of r-f pulses. The two-stage video amplifier has a gain of about 100 which is constant to ±3 db between 200 cps and 4.5 Mc/sec. With this unit installed, the synchroscope is given the designation P4-E. A four-stage wide-band video amplifier may also be installed. This unit has a gain of 2000 which is constant to ±3 db between 120 cps and 5.4 Mc/sec. At the minimum setting of the gain control, the gain is reduced to 200 and the bandwidth is extended to 8.2 Mc/sec. Signal amplitudes of from +0.16 to −0.8 volt may be amplified without overloading. The input impedance is 1000 ohms paralleled by 21 μμf.

No signal attenuator or probe is included with this instrument. It is often used with an external capacity divider for viewing high-voltage pulses, however. Several designs of r-f probes are used with the P4E and are described in Vol. 11 of this series. External mixers are required for viewing several signals simultaneously.

A Wien bridge stabilized oscillator, used to control the *repetition rate* of the trigger generator to ±10 per cent of the frequencies 500, 1000, 2000, and 4000 cps, may be used as a *timing standard*. An oscillator output (sinusoidal waveform) connector is provided. Alternately, the output of a shocked *LC*-circuit having a resonant frequency of 2 Mc/sec may be connected to the CRT deflecting plate by the signal switch. The train of damped oscillations is sustained for a long enough period to allow calibration of the two fastest sweeps. The frequency is accurate to ±5 per cent. A secondary vertical centering potentiometer and meter terminals for reading the voltage across it provide an *amplitude standard.* Pulse amplitude measurements may be made to 2 per cent if a meter of higher accuracy is used. Space is provided for mounting a 3-in. voltmeter on the panel of the synchroscope, and some units are so equipped.

The over-all dimensions of the synchroscope are $8\frac{3}{4}$ by $14\frac{1}{4}$ by $20\frac{1}{2}$ in., neglecting the handle, and the weight is 40 lb. It requires 115 ± 10 volts

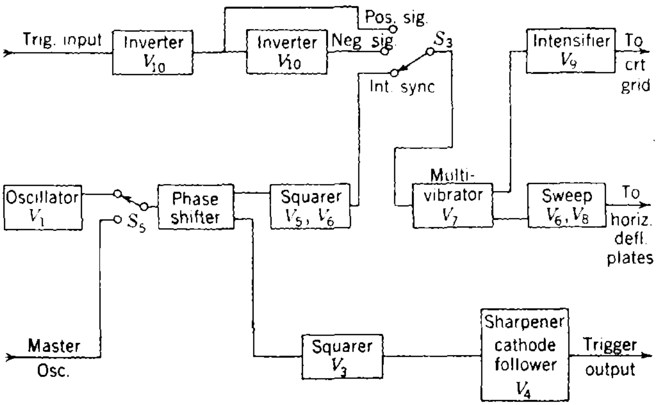


FIG. 18·2.—Block diagram of the P4 synchroscope.

at 50 to 400 cps and has a power rating of 140 watts at 60 cps. The tube complement is 1-5LP1, 2-2X2/879, 1-5Z3, 1-6SK7, 2-6SL7, 6-6SN7, and 1-7V7. The connectors on the instrument are Amphenol type 80C, but no cables are provided. A type "N" r-f connector is used for the r-f signal input the P4E synchroscope. All signal connections are made at the rear of the instrument to reduce lead lengths.

The mechanical construction of this unit is quite conventional as shown in Fig. 18·1. It is built to good commercial standards and is satisfactory for laboratory use.

Circuit Description.—The *cathode-ray tube control circuits* use a high-voltage bleeder of conventional design. Horizontal and vertical centering potentiometers have their range extended below ground by being included with the focus and intensity potentiometers in this bleeder circuit. Application of centering voltages to only one of each pair of deflecting plates results in some defocusing with changes of these voltages. The time base and the video signal are also introduced on only one of their respective pairs of deflecting plates resulting in excessive astigmatism of the trace with large signals. Since this instrument was not designed for precision measurements of observed signals, this was not considered a great disadvantage.

The d-c potential of the signal deflecting plate may be adjusted by the potentiometer whose arm is connected through switch S_4 when this switch is in the ON

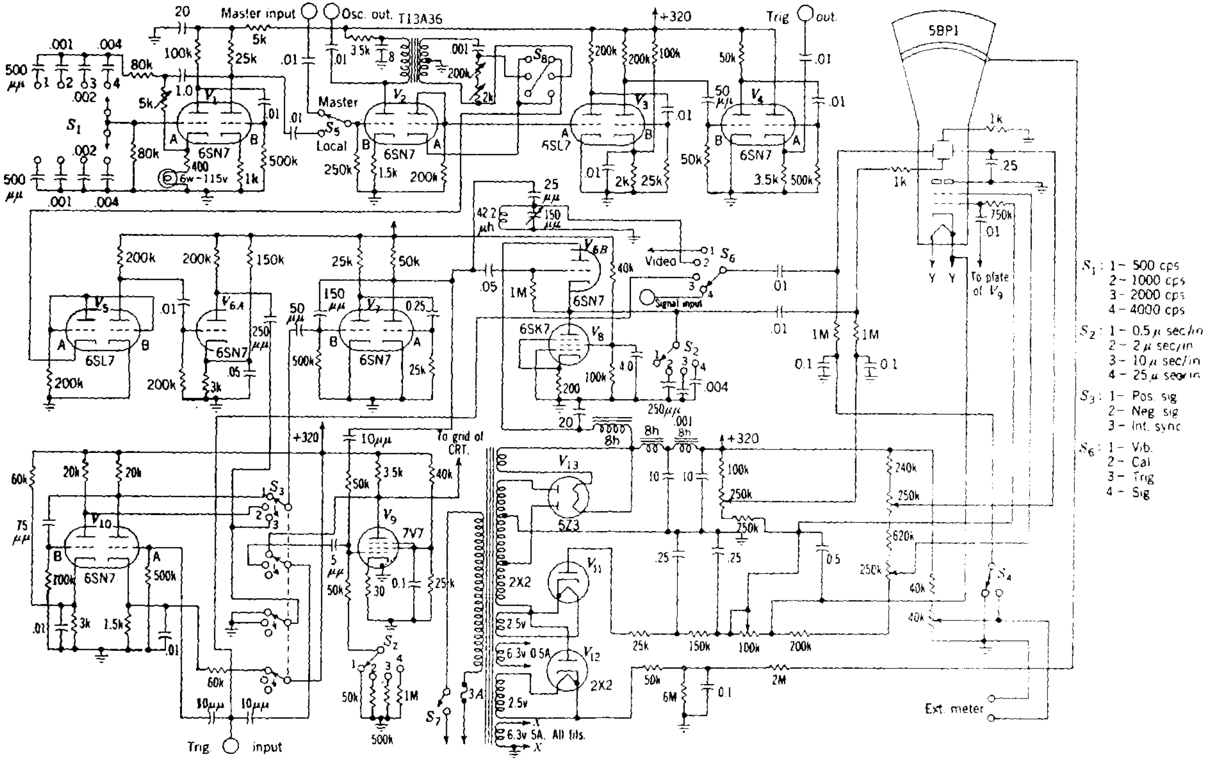


FIG. 18·3.—P4 synchroscope circuit.

position. This permits the measurement of peak voltage by deflecting a signal through its amplitude with an equivalent d-c voltage which may be read directly on an external indicating meter. A switch S_6 is mounted on the back of the instrument and selects signals from one of four locations: the signal input, the trigger input, the calibration oscillator, or the video amplifier (if used).

The *sweep-voltage generating circuits* include a multivibrator, a switch or "clamp" tube, and a constant-current pentode through which the sweep condenser discharges. Constants of the multivibrator have been selected to give a free-running frequency of about 200 cps and to give a maximum rate of rise on the leading edge of the negative half cycle. It may be readily synchronized by negative triggers having a pulse repetition rate between 200 and 4000 cps. These triggers may be obtained from an external trigger generator, being peaked and limited in one or both of the stages of the amplifier V_{10} or from the internal trigger generator as selected by switch S_3. Negative pulses from the multivibrator are coupled to the clamp tube V_{6B}, the intensifying pulse amplifier V_9, and the shocked oscillator composed of the parallel 150-$\mu\mu$f condenser and 42.2-μh choke.

These negative pulses cut off current flow in the clamp tube, permitting the sweep condenser selected by switch S_2 to discharge through the pentode V_8 at essentially constant current. The linear fall of voltage produced is used to deflect the electron beam across the screen of the cathode-ray tube.

The negative multivibrator pulse is also coupled to the control grid of the intensifying pulse amplifier V_9. The pulse rise time of about 0.2 μsec is reduced to a maximum of 0.05 μsec by amplification and by mixing with the differentiated leading edge of the external trigger or signal. Nevertheless, the portion of the leading edge of the pulse that occurs during this period cannot be viewed when the synchroscope must be triggered from an external source. Adjustable phasing between the synchroscope output trigger and the sweep trigger eliminates this difficulty when the system supplying the signal will accept a trigger.

The internal trigger generator consists of a Wien bridge sinusoidal oscillator and amplifier, an adjustable RC phase shifter, and separate squaring and differentiating amplifiers for the output trigger and sweep trigger. An oscillator input connector which permits synchronization from an external sinusoidal voltage of about 15 volts amplitude is also provided.

The frequency of the Wien bridge circuit using the double triode V_1 is set to 500, 1000, 2000, or 4000 cps by the selection of the appropriate pair of condensers with switch S_1. A 6-watt, 115-volt lamp is used as a nonlinear resistance in the oscillator cathode circuit (one arm of the bridge) to provide stabilization of the oscillator amplitude. Another arm, a 5000-ohm variable resistor, may be adjusted to control the waveform.

The oscillator output voltage is amplified by the triode section V_{2B}, and push-pull output is obtained from the transformer in the plate circuit. One winding is connected directly to the reversing switch S_8. The voltage from the other winding has its phase shifted by the RC-divider across the transformer secondary before reaching the switch. These two voltages are then rectified in the diode-connected tube sections V_{2A} and V_{5A}.

One half-wave rectified voltage is then squared in the amplifier tubes V_{5B} and V_{6A} and differentiated by the *RC* in the grid circuit of V_7 to give the internally generated sweep trigger. The other half-wave rectified voltage is similarly amplified by V_{3A} and V_{3B}. This voltage is differentiated by the *RC* in the grid circuit of V_{4B}, and the resulting trigger is amplified by this stage. The cathode follower V_{4A} gives a low-impedance positive output trigger. The reversing

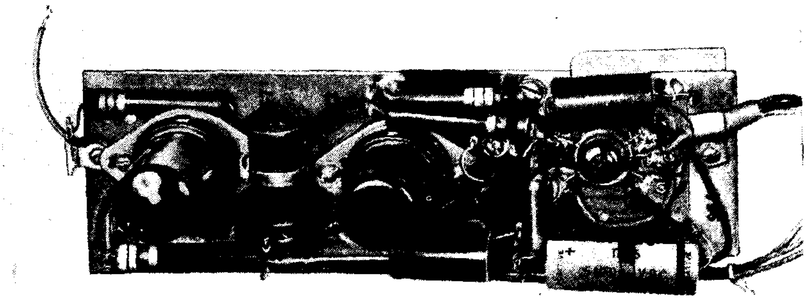


FIG. 18·4.—P4E r-f detector and video amplifier.

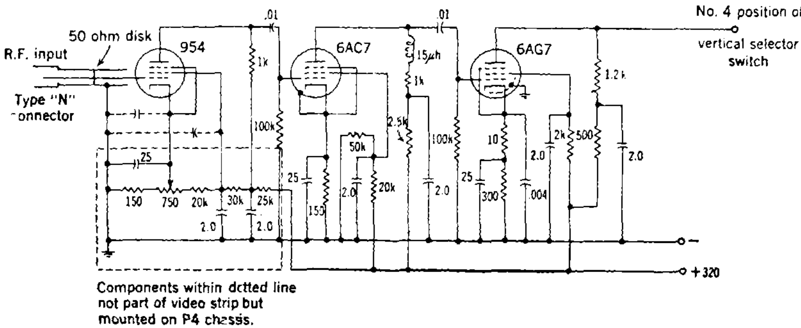


FIG. 18·5.—P4E amplifier circuit.

switch S_8 permits the output trigger to be either advanced or retarded with respect to the synchroscope sweep.

A *signal amplifier* was not included in the P4 synchroscope because it was designed for use with equipment producing pulses of at least 100 volts amplitude. Subsequent application of this unit to small signal and r-f measurements led to the development of several signal amplifiers which can be mounted in the scope at the rear of the chassis.

The circuit of one of these which adapts this instrument for use as an r-f envelope detector is shown in Fig. 18·5. The r-f signal at X band or lower frequency is coupled through a type "N" matched impedance r-f connector paralleled by a 50-ohm disk-type terminating resistor to the control grid of a type

954 acorn pentode. This tube operates as a square law detector, giving an output voltage proportional to the power level of the r-f signal. Gain control is effected by changing the bias level of this tube with the potentiometer in the cathode circuit.

The detector output is amplified in the two-stage video amplifier that follows. The first stage, a 6AC7, uses a shunt peaking coil to increase the plate-load impedance at high frequencies, extending the video bandwidth. A 6AG7 is

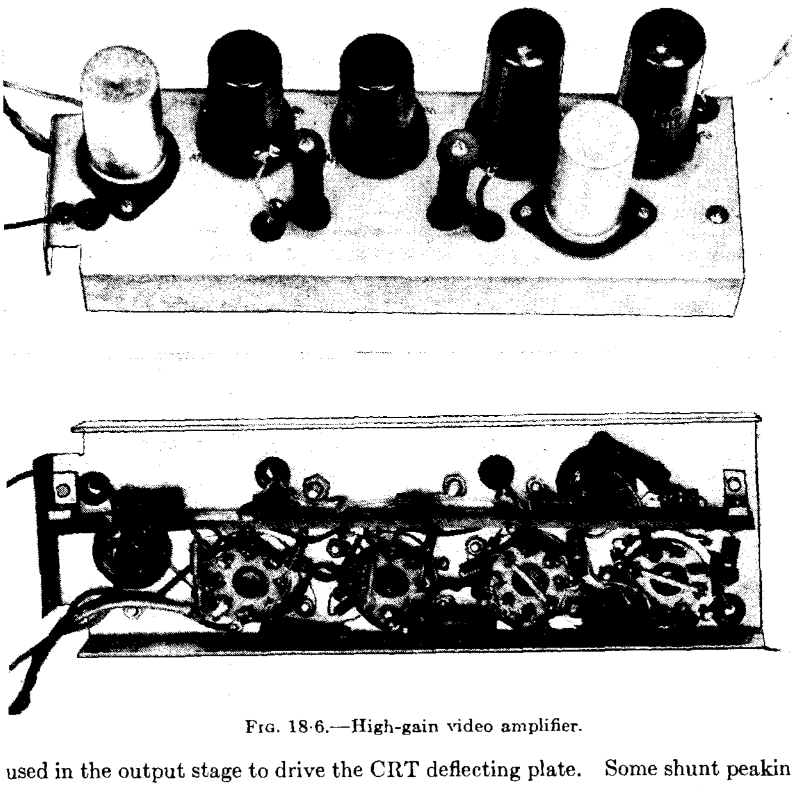


FIG. 18·6.—High-gain video amplifier.

used in the output stage to drive the CRT deflecting plate. Some shunt peaking is provided by the inductance of the wire-wound plate-load resistance in its plate circuit. Further peaking in this stage is obtained by using a very small bypass condenser across the degenerative cathode resistor.

The coupling circuits are designed to pass a wavefront having a rise time of 0.1 μsec between the 5 and 95 per cent points. Since the amplifier is essentially linear, the CRT beam deflection is proportional to the output of the detector and to the power in the r-f input pulse.

The wide-band video amplifier shown in Figs. 18·6 and 18·7 may also be used in this synchroscope when greater sensitivity is required. Shunt peaking is employed in all stages, and the amplifier gain is controlled varying the degenerative cathode resistance in the first stage. This amplifier may also be used as an r-f envelope detector by connecting a probe containing an r-f crystal to the input with not over 12 in. of video cable. With the video gain at the medium setting, the output voltage is nearly proportional to the power input. The electrical characteristics of both of these amplifiers have been described in the *characteristics* part of this section.

A calibrating oscillator is built into this instrument for use as a *timing standard*. The negative pulse from the multivibrator in the sweep generator is coupled to the

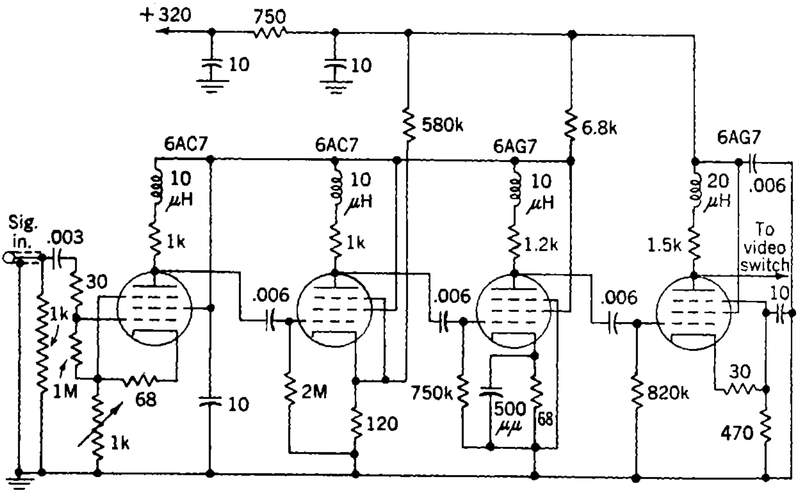


FIG. 18·7.—High-gain video amplifier circuit.

parallel resonant circuit composed of a 150-$\mu\mu$f condenser and 42.2-μh choke. The voltage generated across the circuit is a damped sinusoid having a frequency of 2 Mc/sec. This voltage may be coupled to the vertical deflecting plate of the cathode-ray tube by switch S_6 and is sustained for a long enough period to calibrate the two fastest sweeps. An *amplitude standard* is provided by a circuit for measuring pulse voltages by substitution as previously described.

A *power supply* of conventional design is used. The low-voltage supply gives 320 volts at 115 ma, and both plus and minus 1500-volt cathode-ray tube supplies are included. A separate low-resistance filter for the sweep generator makes 350 volts available for that circuit.

18·2. The Model 5 Synchroscope. *Function.*[1]—The Model 5 synchroscope is a laboratory instrument designed primarily for determining the

[1] The Model 5 synchroscope was designed by the MIT Radiation Laboratory and manufactured by Sylvania Electric Products, Inc., and Philharmonic Radio Corp. Procurement was by the Radiation Laboratory and others. References are RL Reports Nos. M-212 and S-18.

amplitude, duration, and shape of short video pulses. Its functions are quite similar to those described for the P4 synchroscope, but its circuits have a much greater stability, so that more precise measurements can be made. While it is primarily intended for use in the study of radar modulator and magnetron pulses, it has found wide application in the design and production testing of all types of pulse circuits in radar and allied fields. Its construction for the most part meets Army and Navy

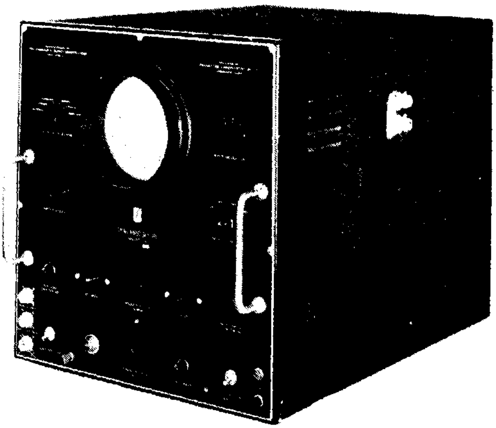


FIG. 18·8.—Model 5 synchroscope.

specifications, but its portability is considerably restricted by its weight and bulk.

Characteristics.—A type 5JPI *cathode-ray tube* operating at a total accelerating potential of 2700 volts is used in this instrument, giving a vertical deflection factor of approximately 65 volts per in. This tube was chosen because of its very low shunting capacity (about 3 $\mu\mu$f) at the deflecting plates.

A triggered *linear-sweep voltage generator* of the feedback type is used. The horizontal time base writing speeds are 0.01, 0.05, 0.2, 1, 2, and 5 in./μsec with a minimum sweep amplitude of 4 in. on the 5-in. cathode-ray tube screen. Either the output of the internal trigger generator or an externally generated trigger or sine wave having a frequency of 50 to 5000 cps may fire the sweep generator. For externally triggered operation a positive or negative trigger having a minimum amplitude of 50 volts and a rate of rise of 70 volts/μsec is required. Alternatively, a sinusoidal voltage having a peak-to-peak amplitude of at least 30 volts may be used. Sweep voltage input or output connectors are not provided on this instrument.

The output of the *internal trigger generator* is a positive trigger which has an amplitude of 200 volts, repetition rates of 500, 1000, 2000, and 4000 cps and which is supplied from a low-impedance source. This pulse has a duration of 15 μsec between 10 per cent amplitude points and rises to full amplitude in 0.3 μsec. It is phasable from 75 μsec before to 25 μsec after the start of the sweep generator.

The standard *signal input circuits* comprise six input fittings which may be connected to the top vertical deflecting plate through a selector switch. The input impedance is 470,000 ohms, paralleled by 15 μμf.

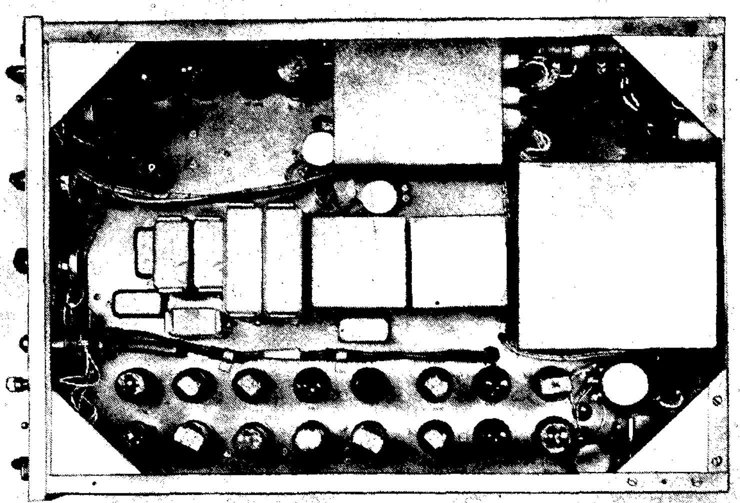


FIG. 18·9.—Model 5 synchroscope, bottom view.

A *video amplifier* is not normally supplied with this instrument. Space and power are provided, however, for either of the amplifiers described for use with the P4 synchroscope. An amplifier of more recent design may be mounted in this unit if an external auxiliary power supply is used. It has a gain of approximately 180 and is flat within 3 db between 20 cps and 18 Mc/sec. The input impedance is 1 megohm paralleled by 14 μμf. A six-step resistance-capacity attenuator having signal attenuation ratios of 1, 3, 10, 30, 100, and 330 to 1 is included in the input circuit. The maximum unattenuated input voltage that can be amplified without overloading is 0.3 volt. This may be increased by the attenuation ratio at other attenuator settings. No probes or signal mixing facilities are provided.

No *timing standard* is incorporated in this instrument. Sweep calibration and the measurement of time intervals are accomplished with the aid of separate marker generators. An *amplitude standard* is provided by an auxiliary vertical centering control which is used to measure pulse amplitudes by substitution. The voltage readings are made on an external meter which connects to pin jacks mounted on the panel. The over-all dimensions of this unit are 22 by 15½ by 17 in.; the weight without auxiliary amplifiers is 85 lb. A supply voltage of 115 ± 10 volts, 60 cps is required at a power drain of 200 watts. The tube com-

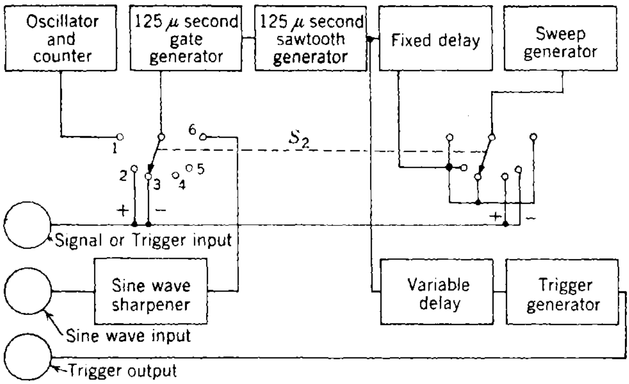


FIG. 18·10.—Block diagram of the Model 5 synchroscope.

plement is 1-5JPI, 2-2X2, 1-5U4G, 3-6AG7, 1-6H6, 2-6SH7, 10-6SN7, 1-6X5, 1-6Y6, 1-VR105, and 1-VR150. All cable fittings are of the UHF single-conductor type with the exception of one Amphenol type 80-PC 2F two-conductor fitting on the rear of the unit. The *mechanical construction* of this unit departs from the conventional in mounting tubes below and components above the chassis. While it does not entirely meet wartime standards, it is well built and should be satisfactory for field as well as laboratory use.

Circuit Description.—The circuits that supply the operating potentials to the *cathode-ray tube* elements have been carefully designed to give optimum focusing over the entire screen. A well-filtered negative potential of 1300 volts supplies the high-voltage bleeder that contains the focus and intensity controls. An after-acceleration potential of 1500 volts is also included. Balanced horizontal centering voltages are obtained from a dual potentiometer. Another dual potentiometer is used to vary the average horizontal centering potential and the second anode potential by equal amounts. This astigmatism control is screwdriver adjusted and is mounted through the chassis.

Single-plate vertical centering is necessitated by the inclusion of substitution voltage measuring circuits. The vertical centering control varies the potential of the lower deflecting plate, while the potential of the upper vertical deflecting

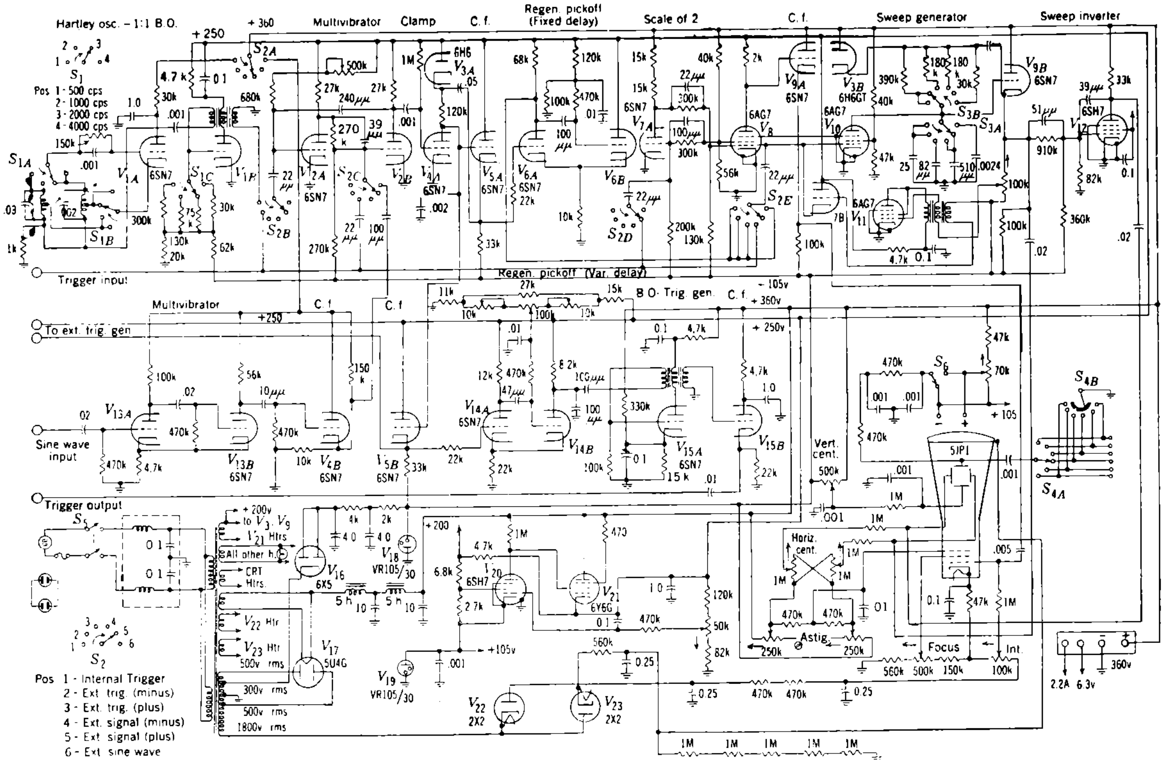


FIG. 18·11.—Model 5 synchroscope circuit.

plate may be adjusted by the CALIBRATE control. In making pulse amplitude measurements, the latter control is used to deflect a pulse through a distance equal to its amplitude. The pulse amplitude can then be read directly on a 1000-ohm per volt external meter which is connected to the VM tip jacks on the panel. Any of six signal input jacks may be connected to this upper deflecting plate by the INPUT SELECTOR switch S_4. The *linear-sweep generator* consists of a scale-of-two type of multivibrator, a bootstrap-type sweep generator, a sweep inverter stage, and a circuit that terminates the sweep on all ranges when the proper voltage amplitude is reached. The multivibrator made up of V_{7A} and V_8 has two stable states and regenerates from one to the other with each trigger applied. A trigger from either the internal or external trigger generator fires V_{7A} and cuts off V_8. This condition is maintained until the sweep voltage attains the required amplitude, at which time a positive trigger from the cathode follower V_{7B} again fires V_8. This restores the initial condition until the next input trigger is received.

The positive voltage pulse produced at the plate of V_8 is coupled to the cathode follower V_{9A}, whose output intensifies the cathode-ray tube during the sweep. The cathode follower reduces the shunt capacity at the plate of V_8, decreasing the rise time of this waveform. When the synchroscope is triggered by the signal to be viewed, it is essential that the sweep and intensifying pulse start as rapidly as possible. Only about 0.01 μsec is lost in viewing a fast-rising negative signal, although slightly more time is lost when a positive signal triggers this circuit.

The fall of voltage at the grid of V_8 is direct-coupled to the grid of the clamp tube V_{10}. This tube is cut off, allowing the sweep condenser selected by switch section S_{3A} to charge through the charging resistor selected by switch S_{3B}. The voltage across the charging resistor is maintained essentially constant by the bootstrap cathode follower V_{9B}, thereby generating a linear rise of voltage across the sweep condenser. The phase of this signal is reversed by the voltage-feedback inverter V_{12}, and both signals are capacitively coupled to the horizontal deflecting plates.

When the voltage at the arm of the potentiometer in the cathode circuit of V_{9B} rises above the cutoff potential of the blocking oscillator V_{11}, regeneration in this stage occurs. The positive pulse generated across the tertiary winding of the pulse transformer is coupled to the cathode follower V_{7B}. As previously mentioned, this drives the grid of V_8 (in the multivibrator) positive, terminating the gate and the sweep. This eliminates any requirement for switching time constants in the multivibrator and maintains a fairly constant sweep length on the cathode-ray tube screen. This length is normally adjusted to 4.5 in. with the 100-k screwdriver-controlled potentiometer which sets the control grid potential of V_{11}.

The *internal trigger circuits* include a Hartley oscillator, a blocking oscillator trigger generator, and a linear-sweep time-modulation circuit having two "pick-off" stages. These make possible several types of operation which are chosen by the TRIGGER SELECTOR switch S_2. In one of these, the action of the delay circuit may be initiated by the internal trigger, an externally generated positive or negative trigger or a sinusoidal voltage. The sweep trigger is generated after a fixed delay of 90 μsec, and the positive output trigger may be delayed by an amount adjustable from 75 μsec before to 25 μsec after the sweep trigger. Alternatively, the sweep generator may be triggered directly by the video signal.

The internal trigger generator consists of the *LC*-tuned oscillator V_{1A} and the blocking oscillator trigger generator V_{1B}. The tank circuits selected by the REPETITION RATE switch S_{1A} are tuned to 1000 and 4000 cps. The blocking oscillator is synchronized by the voltage pulses developed across the 1000-ohm resistor in the cathode circuit of V_{1A} by the pulses of current resulting from Class C operation. The grid time constant of the blocking oscillator is changed by switch S_{1C} permitting the generation of 500 and 2000 cps repetition rates by frequency division.

A linear-sweep delay circuit having dual pick-off stages supplies the sweep and output triggers. The multivibrator made up of V_{2A} and V_{2B} is fired by the pulse selected by the TRIGGER SELECTOR switch S_2. The negative gate generated at the plate of V_{2B} is coupled to the grid of the clamp tube V_{4A}, cutting it off. The charging of the sweep condenser through the plate resistor of V_{4A} is made linear by the "bootstrap" cathode follower V_{5A}.

After the cathode potential of V_{5A} has risen for a fixed period of time, the grid of V_{6A} is carried past its cutoff potential. This tube and V_{6B} are used in a cathode-coupled multivibrator circuit which immediately regenerates. The leading edge of the negative gate produced at the plate of V_{6A} is differentiated to generate the sweep trigger.

The variable delay pick-off stage is isolated from the delay sweep generator by the cathode follower V_{5B}. The pick-off multivibrator V_{14} is identical with the one just described except that the cutoff potential of the input tube may be varied with the TRIGGER-DELAY COARSE and FINE controls, which set the quiescent common cathode potential by adjusting the voltage to which the grid of the "on" tube is returned. The differentiated rise of the positive gate from this stage triggers the blocking oscillator V_{15A}, which supplies the positive output trigger through the cathode follower V_{15B}.

When a sinusoidal input voltage is used for synchronization, it synchronizes the cathode-coupled multivibrator V_{13}. The square waves developed at the plate of V_{13B} are differentiated, and the negative pulses removed in the cathode follower V_{4B}, which supplies a positive trigger to the delay circuits.

This unit is not usually supplied with a *video amplifier* installed. Space and power are available, however, for either of the amplifiers described for use with the P4 synchroscope (Sec. 18·1). The schematic diagram of a five-stage video amplifier is shown in Fig. 18·14. This amplifier takes advantage of the low interplate capacity of the 5JPI and the high-power output of the 829 push-pull amplifier to give a flat frequency response (±3 db) between 20 cps and 18 Mc/sec. Details of construction and mounting are shown in the accompanying photographs. An input attenuator of the resistance-capacitance divider type is used as a gain control. Included in the amplifier proper are two 6AC7 amplifier stages and a 6AC7 paraphase inverter-amplifier. These drive a pair of push-pull 6AG7's, which in turn drive the two sections of the 829 in push-pull. Both shunt and series peaking are used in all stages except the inverter, in which only series

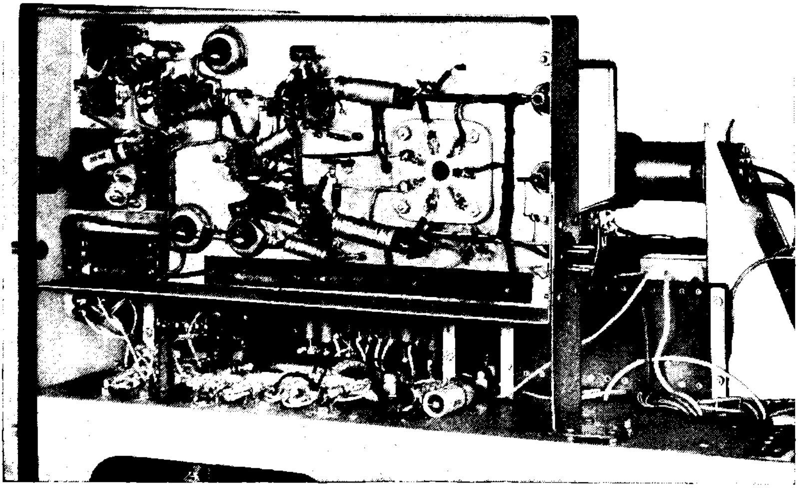


FIG. 18·12.—Bottom view of the wide-band video amplifier.

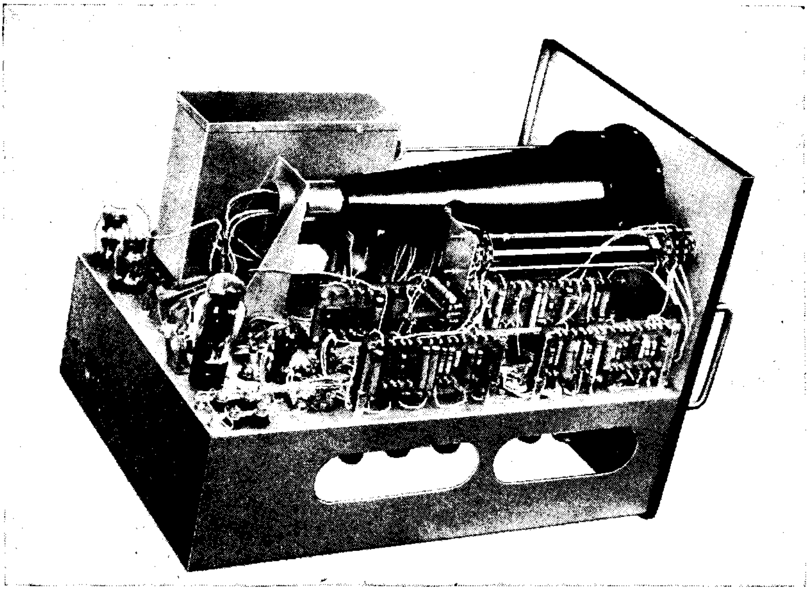


FIG. 18·13.—Model 5 synchroscope without wideband amplifier, inside view.

peaking is employed. Stray capacities throughout the amplifier have been kept to a minimum by careful layout and point-to-point wiring. The sizes of the

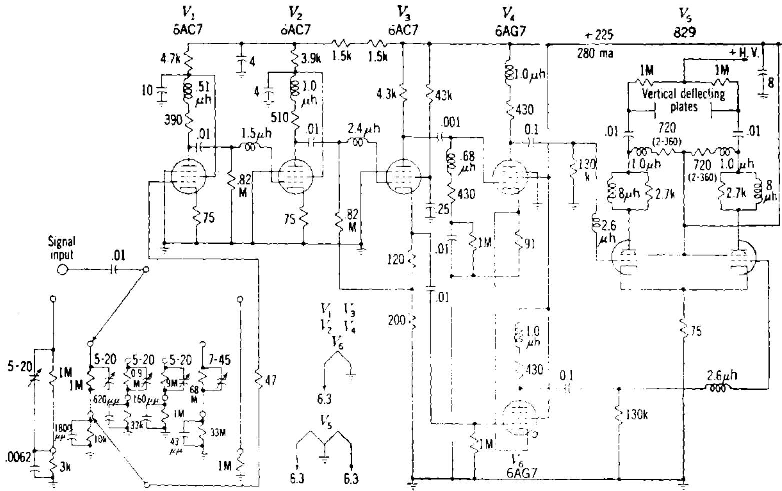


FIG. 18·14. Wide-band video amplifier circuit.

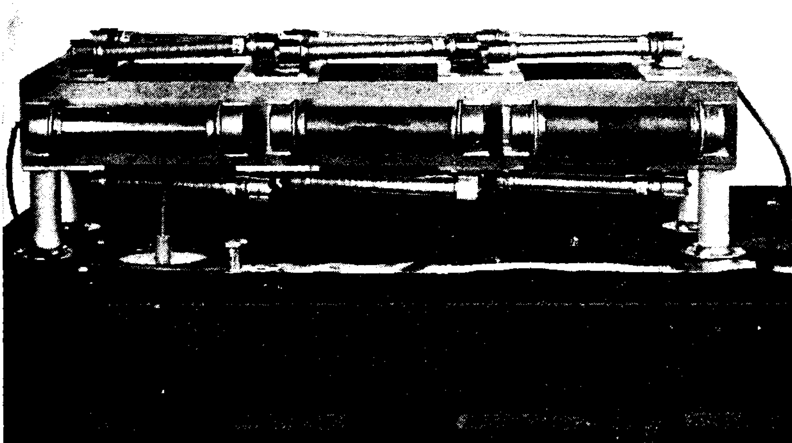


FIG. 18·15.—High-voltage multiplier.

peaking inductances are correct for the particular layout shown but will require readjustment for any changes in components or layout.

An external power supply delivering 225 volts direct current at 280 ma and 6.3 volts alternating current at 7 amp is required for this amplifier. It is of conventional design and uses a pair of 5R4-GY rectifier tubes. The *power supply* delivers +1500, −1300, and +360 volts unregulated in addition to regulated potentials of +250 and −105 volts. The 250-volt potential may be set by a potentiometer controlling the control grid voltage of V_{20}. The use of this electronically regulated supply is essential if "jitter" in the delay and sweep circuits used is to be avoided.

A voltage multiplier to supply 12,000-volt accelerating potential for a type 5RPI cathode-ray tube is shown in Figs. 18·15 and 18·16. Sufficient brilliance is obtained for direct screen photography of the very fast sweeps available in this unit.

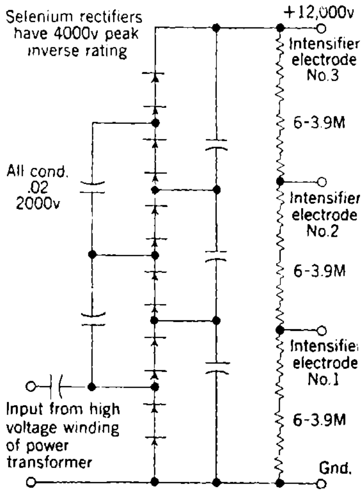


FIG. 18·16.—High-voltage multiplier circuit.

18·3. Oscilloscope TS34/AP.[1]

Function.—The TS34/AP test oscilloscope shown in Fig. 18·17 is a small, easily portable instrument that was designed for use as a general-purpose oscilloscope for laboratory and field maintenance of radar and similar electronic equipment. It is useful for signal tracing and for making voltage and power measurements by comparison.

Characteristics.—A type 2API cathode-ray tube operating at an accelerating potential of 800 volts is used. A viewing hood is built into the oscilloscope with a lens over the face of the tube to give an image magnification of approximately two.

Both triggered and free-running sawtooth *sweep generators* are provided. The triggered sweep lengths are 5, 50, and 250 μsec, while the duration of the free-running sweep is variable from 20 to 100,000 μsec. Triggered sweep rates of 10 to 1800 cps may be used with the 250-μsec sweep triggered by all synchronizing pulses. Higher rates may be used by frequency dividing in the sweep multivibrator. Free-running sawtooth sweep rates of 10 to 50,000 cps are provided. Both the triggered and sawtooth sweep generators may be synchronized by external signals of either polarity having amplitudes between 0.5 and 75 volts and by observed signals of 0.1 volt or greater. Positive and negative sweep voltages of 150 volts amplitude can be supplied to a load impedance of

[1] Designed by the Bell Telephone Laboratories and manufactured by the Western Electric Co., Code No. X-61713-B. Instruction manuals include TM 11-1067 and AN 08-35 TS34-2 published by the Dayton Procurement Office and the manuscript of the *Handbook of Maintenance Instructions for Oscilloscope TS34/AP* published by Western Electric Co. TS34A/AP is an improved version having continuously variable sweep circuits and being drip-proof and vibration-proof.

not less than 500,000 ohms from an output jack on the instrument. Also, a single-polarity sweep of 300 volts amplitude or a push-pull sweep of ±150 volts amplitude may be supplied to the oscilloscope by an external generator.

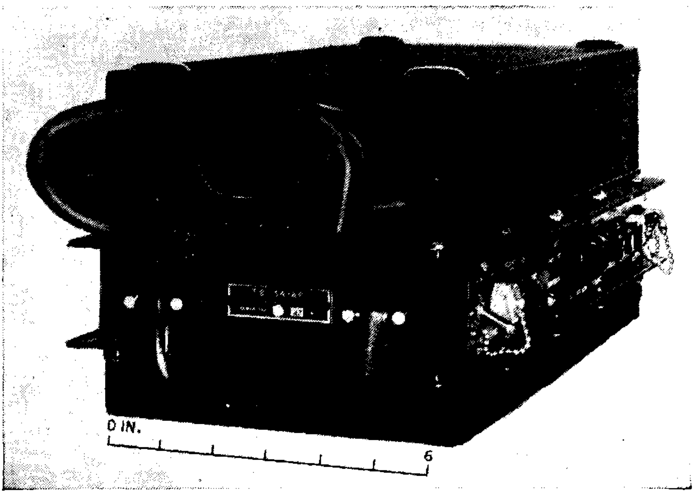


FIG. 18·17.—Oscilloscope TS34/AP.

The TS34/AP was designed to operate only with radar systems or portions of systems that supply a synchronizing trigger. Consequently an internal *trigger generator* is not incorporated in the design.

A direct *signal channel* to one deflecting plate has an input impedance of 3.9 megohms paralleled by 25 $\mu\mu$f. Also included is a video amplifier which has a gain of about 1000, constant to ±3 db between 40 cycles and 3 Mc. A 0.5-μsec delay line is included in this circuit to permit starting the sweep before the signal reaches the deflecting plate. With no attenuation, a signal voltage of 0.1 to 1 volt may be amplified without overloading. This limit is increased to 100 volts on the 40-db input position and to 450 volts when the probe is used. The input impedance on the low-impedance position is 62 ohms and on the high impedance position is 430,000 ohms paralleled by 30 $\mu\mu$f. A signal attenuator covering the range of 0 to 60 db in 2-db steps and a signal probe of the resistance-capacitance divider type are included in the design. The probe gives a voltage attenuation of 10/1 and has an input impedance of 4 megohms paralleled by 12 $\mu\mu$f.

A cathode input permits signal mixing by intensity modulation of the cathode-ray tube beam. A connection to the unused vertical deflecting plate is also available as described above. It should be remembered that an 0.5-μsec time difference exists between these signal inputs and the regular video amplifier channel.

An ingenious *timing standard* is built into this oscilloscope. Provision for improperly terminating the 0.5-μsec video delay line is included. When a fast rising pulse is impressed on the video input, a series of timing pulses are produced by multiple reflection. These may be used for calibration of time intervals not longer than 10 μsec. While the calibrated attenuator may be used for relative voltage measurements, there is no voltage standard in this instrument.

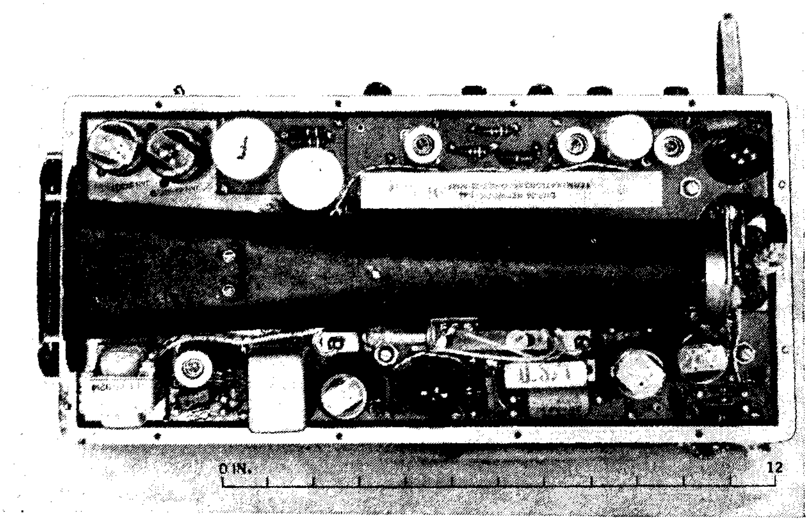


FIG. 18·18.—Top interior view of the TS34/AP.

The dimensions of this oscilloscope are 6 by 8 by 15 in. over all, and the carrying case is 29 by 9¾ by 10 in. Its weight is 26 lb alone or 46 lb when in the carrying case with all cables. Power requirements are 115 ±10 volts, 50 to 1200 cps, and 90 watts at 60 cps. Ambient temperature limits are −40° to +120°F. The tube complement is 1-2AP1, 1-5Y3GT, 1-6AC7, 1-6AG7, 4-6AK5, 2-6SL7GT, 1-6SN7GT, and 1-6X5GT.

The TS34/AP test oscilloscope has been designed to keep weight and volume at a minimum, while taking advantage of the ease of manufacture derived from the use of complete circuit subassemblies. In this unit the subchassis is mounted in a rectangular frame of channel construction. The controls are mounted in this channel and are placed immediately adjacent to their respective circuits. "Dish" type covers are attached to both sides of this rectangular frame by captive-head screws. The cathode-ray tube mounting is a tubular magnetic shield which also serves as a portion of the viewing hood and supports the magnifying lens and illuminated scale.

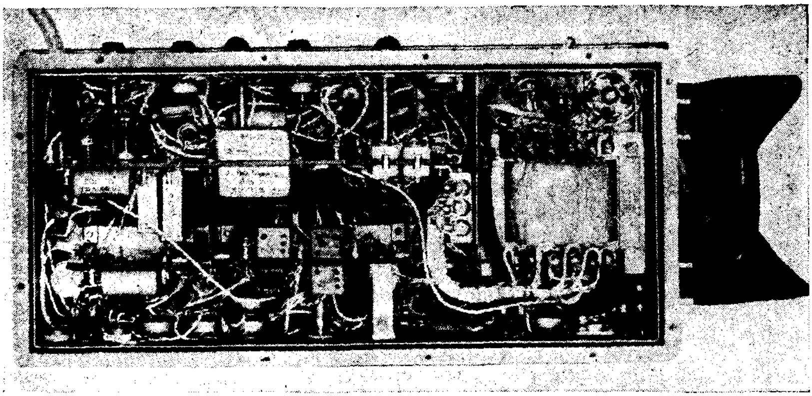


FIG. 18·19.—Bottom interior view of the TS34/AP.

Circuit Description.—The *cathode-ray tube control circuits* have a conventional high-voltage bleeder circuit with potentiometers for focus and intensity control. Vertical centering voltage is applied to one deflecting plate only, the opposite plate having a d-c connection to the output of the video amplifier. Horizontal centering is accomplished by changing the d-c level of the bypassed grid of the cathode-coupled sweep amplifier. When switch S_6 is opened, a 2.2-megohm resistance is introduced into each of the leads from the sweep amplifier to the horizontal deflecting plates, and external sweep voltages may be introduced through the connectors marked H-1 and H-2. The connector marked TIME may be used to apply signals to the cathode of the cathode-ray tube, giving intensity modulation. This circuit has a short time constant (R = 5100 ohms; C = 100 $\mu\mu$f) and is most suitable for fast timing markers.

The synchronizing pulse amplifier V_{10} is used with both the triggered and sawtooth *sweep generators*. A synchronizing signal input, which may be either positive or negative, can be obtained from the video signal after the first amplifier stage or from an external source. Variation of the cathode degeneration controls the stage gain and the synchronizing signal amplitude supplied to the sweep generator.

A multivibrator type of circuit having one stable condition is used to generate the triggered sweep voltage. One triode section of V_8 is the normally nonconducting tube, while the cathode, control grid, and screen grid of V_9 act as the normally conducting triode of the multivibrator. Current flowing in the plate circuit of V_9 holds the quiescent voltage to ground of that element to about 50 volts. When the negative synchronizing pulse triggers the multivibrator, the current flow in V_9 is cut off and the voltage across the sweep condenser con-

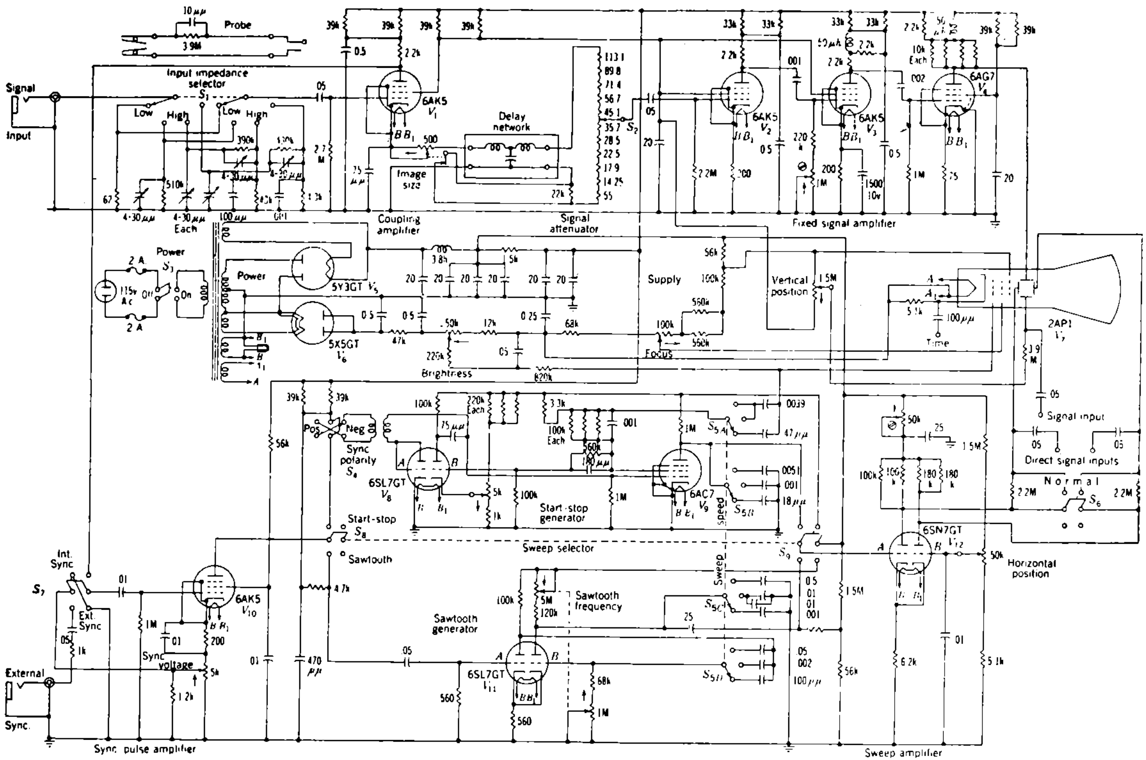


FIG. 18·20.—TS34/AP circuit.

nected between plate and ground rises exponentially toward the supply potential. Reasonable linearity is obtained by amplifying only the first 5 per cent of this exponential to produce the push-pull sweep voltage. The other section of V_8 is used as a diode clipper to eliminate positive pulses and overshoots present in the synchronizing signal. It also prevents the control grid of V_9 from rising above ground, holding the quiescent plate voltage constant with variations of duty cycle. The 5000-ohm potentiometer in the cathode circuit of V_{8B} is adjusted to just prevent continuous oscillation of the multivibrator, giving maximum sensitivity to synchronizing signals.

The sawtooth sweep voltage is generated by a continuously oscillating, synchronized multivibrator using the two triode sections of V_{11} in a cathode-

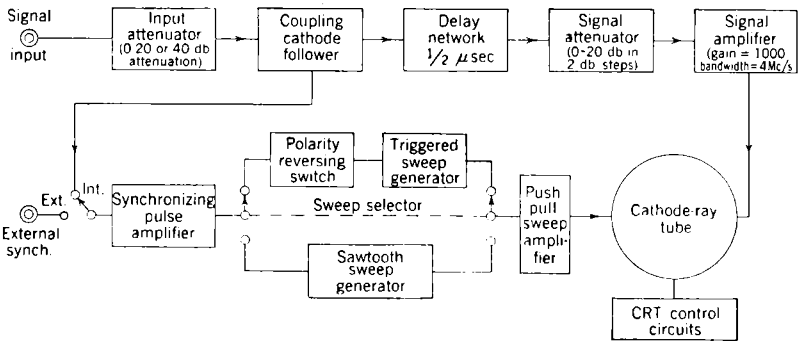


FIG. 18·21.—Block diagram of the TS34/AP.

coupled type of circuit. Generation of the sweep voltage is accomplished by a partial charge and discharge of the sweep condenser connected between one plate and ground with each cycle of oscillation. The time constant of this circuit is adjusted to keep the voltage change small with respect to the supply voltage for all sweep frequencies. This gives a reasonably linear sawtooth which is amplified in the sweep amplifier. Coarse frequency control is accomplished by changing the capacity of the coupling condenser between the plate of one tube section and the grid of the other with switch S_{5D}. The fine frequency control varies the discharge time constant of this coupling condenser. The potentiometer in the plate circuit of V_{11B} is in the charging circuit of the sweep condenser selected by switch S_{6C} and keeps the time constant of that circuit proportional to the time constant of the grid circuit so that a constant sweep voltage output is obtained.

Approximately balanced push-pull output voltages are obtained from a sweep amplifier of the cathode-coupled type. Since these voltages are direct-coupled to the cathode-ray tube deflecting plates, a control has been provided to adjust the average potential of the two plates to minimize astigmatism. The horizontal centering control adjusts the d-c level of the bypassed grid changing the potential differences between the plates.

The input attenuator provided in the *signal amplifier channel* of this oscilloscope is of the resistance-capacitance divider type described in Chap. 17. Trimmer capacitors which parallel the resistors between the 20- and 40-db pairs of switch points on switch S_1 are adjusted to pass a 20 kc/sec square wave without peaking or rounding off the rising and falling edges. This adjustment will be correct when the ratio of the trimmer capacity to the grid-to-ground capacity equals the ratio of the grid-to-ground resistance to the resistance across the trimmer. The capacitors between these switch points and ground are adjusted in a similar manner when the probe is in use.

The coupling amplifier V_1 drives the 12-section 0.5-μsec video delay line in its cathode circuit and the synchronizing pulse amplifier from its plate circuit. The image size control changes the stage gain by varying the amount of negative feedback.

The delay network is terminated in a calibrated attenuator composed of resistors mounted on switch S_2, and its output drives the three stages of a fixed-gain video amplifier. In the first amplifier V_2, a relatively small bypass condenser across part of the plate-load resistor is used to give some low-frequency compensation. The second amplifier V_3 employs shunt peaking to extend the high-frequency response in addition to the low-frequency compensation. A variable resistor in the grid circuit of V_3 permits adjustment of the over-all low-frequency response of the amplifier. Shunt peaking is employed in the final amplifier stage V_4, which is direct-coupled to one vertical deflecting plate of the cathode-ray tube.

In building this amplifier, it should be remembered that any departure from the original mechanical design may change the operating characteristics materially. More detailed information may be obtained from the maintenance handbook previously mentioned.

The *power-supply* circuit employed in this oscilloscope is entirely conventional in design. If these circuits are to be used with a 3- or 5-in. cathode-ray tube, a higher voltage will be required.

18·4. Type 256-B A/R Range Scope.[1] *Function.*—The A/R range scope[2] (Fig. 18·22) was designed to fulfill the multiple functions of a range circuit calibrator, an auxiliary range unit, and a test scope for radar systems. It contains a crystal-controlled range mark generator, interpolating circuits, sweeps, and a video frequency amplifier suitable for observing fast pulses. The operational flexibility of the unit as a general-purpose test scope has been curtailed to achieve good calibrator accuracy with reasonable portability. Since this unit was designed to measure

[1] Section 18·4 is by H. J. Reed, Jr., and A. H. Fredrick.

[2] Developed by the Radiation Laboratory and manufactured by the Allen B. Du Mont Laboratories, Inc. Procured by the Navy Department and the Radiation Laboratory. Instruction manuals and reports include NAVSHIPS 900, 605 and CO-NAVAER 16-55-504, published by the Bureau of Aeronautics, Navy Department, Washington, D.C., and RL Report No. 755. The schematic diagram is RL drawing number D 1852-A. The type 256-A is similar to the 256-B, the principal difference being an 800-yd "R" sweep on the 200,000-yd. scale of the 256-A instead of the 4500-μsec sweep of the 256-B. A closer approach to a unit meeting service specifications is obtained in the 256-B.

time in terms of radar distances, the "radar yard" (0.0061 μsec) will be used in this discussion.

Characteristics.—A 5CP1 cathode-ray tube operating with 2000 volts on the gun structure and 2000 volts of post-deflection acceleration is

FIG. 18·22.—Type 256-B A/R range scope.

used to obtain definition and brilliance consistent with the sweep speeds and video bandwidth provided.

The *sweep generator* is of the triggered type, producing both type "A" and type "R" displays. Type A sweep lengths are 1220, 122, 24.4, 12.2, and 4.9 μsec (200,000; 20,000; 4,000; 2,000; and 800 yd), corresponding to 300, 30, 7, 3, and 1 μsec/in. of trace. On the type 256-B units, a 4500-μsec (1000 μsec/in.) sweep is also available.

Switch positions that allow the 4000-, 2000-, and 800-yd sweeps to be delayed between 500 and 20,000 yd and the 4000 and 2000 yd sweeps to be delayed from 1000 to 200,000 yd are provided. These delays are controllable through a panel dial which may be calibrated to be accurate to better than 0.1 per cent. The portion of the 20,000- or 200,000-yd A sweep, bracketed by the fast R sweep, can be intensified. Details can thus be picked out without the necessity of an A sweep calibration matching the delay dial.

Repetition rates are limited to a maximum of 2000 cps on the 20,000-yd scale and 400 cps on the 200,000-yd scales. The minimum trigger required for accurate ranging is ± 15 volts with a rate of rise of 100 volts/μsec. Satisfactory triggering can be obtained with rates of rise as low as ± 10 volts/μsec, but recalibration of the sweep delay circuit may be needed. Externally generated input triggers are shaped in an input circuit that supplies an essentially constant trigger to the internal circuits for all trigger amplitudes greater than the minimum required. This shaped trigger is also available on panel connectors. There is no provision for bringing out the sweep voltages or for using externally generated sweep voltages in this instrument.

The start (left edge) of the R sweeps is intentionally made very sharp. The range delay dial, controlling the position of the R sweeps, can thus be accurately calibrated to indicate the elapsed time or range between the trigger and the start of the sweeps. The normal accuracy of this calibration is better than 0.1 per cent of full scale; 20 yd in 20,000 or 200 yd in 200,000. Provision is made for calibrating this delay with the crystal-controlled range marks.

An 81.94-kc/sec crystal oscillator, accurate to 0.02 per cent, is used as a *timing standard*. Marker pulses 1 μsec wide, rising in 0.25 μsec and occurring at intervals of 2000 or 10,000 yd, are derived from this standard. These pulses can be used as bright or dark dot markers on the scope sweeps. Owing to the nature of the trigger generator circuit, 2000 yd are lost in the first 10,000-yd marker interval. Consequently the first "10,000"-yd marker occurs at 8000 yd. Subsequent marks have the correct spacing. Marker spacing and polarities are chosen by a panel switch, and the markers chosen are simultaneously available at an external connector. Pin jacks on the panel bring out the vertical centering voltage for pulse *amplitude measurements* with an external meter.

An internally generated ± 100-volt *output trigger* having a width of 1 μsec and a rise time of 0.3 μsec is also obtained from the divider chain. This trigger is accurately locked in phase with the 2000-yd markers. Its frequency is adjustable to any integral submultiple of 81.94 kc/sec between 80 and 2000 cps on the 20,000-yd range and between 80 and 400 cps on the 200,000-yd and 4500-μsec ranges.

One *signal channel* is a direct connection to the top vertical deflection plate of the 5CP1 and is accessible when the attenuator is set at OFF. The deflection factor for this plate is 79 ± 16 volts per inch upward for a positive signal and the input impedance is 1 megohm paralleled by 32 μμf. Maximum signal level on this channel is 600 volts d-c plus peak a-c.

The lower 5CP1 vertical deflecting plate is accessible through a three-stage *video amplifier* having a gain of approximately 125. The amplifier gain is down 3 db at 7 Mc and 6 db at 10 Mc. Its pulse response is such that the sum of the output rise and fall times does not exceed 0.08

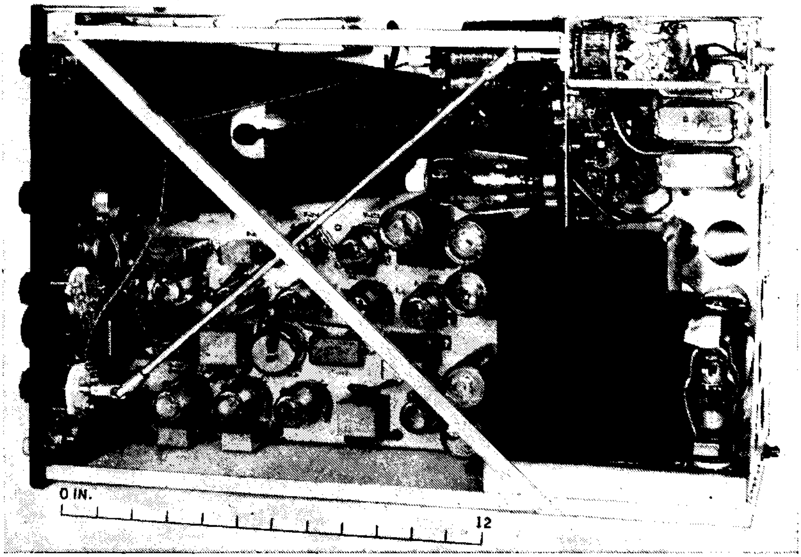


FIG. 18·23.—Right interior view of the A/R range scope.

μsec for an input pulse with a rise and fall time of 0.01 μsec. The low-frequency response is such that the vertical position of the trace following a 1000-μsec pulse does not change by more than 10 per cent of the pulse amplitude.

A 0.2-volt signal at the input of the amplifier gives at least ¾-in. scope deflection. Overload occurs for a 1-volt peak input signal at which point the scope deflection is approximately 1.5 in. The peak signal plus steady voltage at the amplifier input should not exceed 600 volts. A step attenuator having ratios of 1/1, 3/1, 10/1, 30/1, and 100/1 permits the observation of signals in excess of 1-volt amplitude. This attenuator is made up capacity-compensated resistance divider units which are switched in ahead of the video amplifier. The input impedance is approximately 1 megohm shunted by 20 μμf on all ranges.

Internally generated range marks and the R sweep gate are introduced as intensity modulation of the beam. A minor wiring change will permit simultaneous connection to both vertical deflecting plate with one direct channel and one through the amplifier for mixing two video signals.

This equipment is housed in a steel case $11\frac{3}{8}$ by $16\frac{1}{4}$ by 26 in. The case is equipped with two carrying handles, access panels for internal adjustments, and a panel cover containing clips for the instruction manual and power cord. Total weight is 104 lb. Power requirements are 115 to 120 volts, 60 to 1200 cps, 220 watts at 60 cps. The tube complement

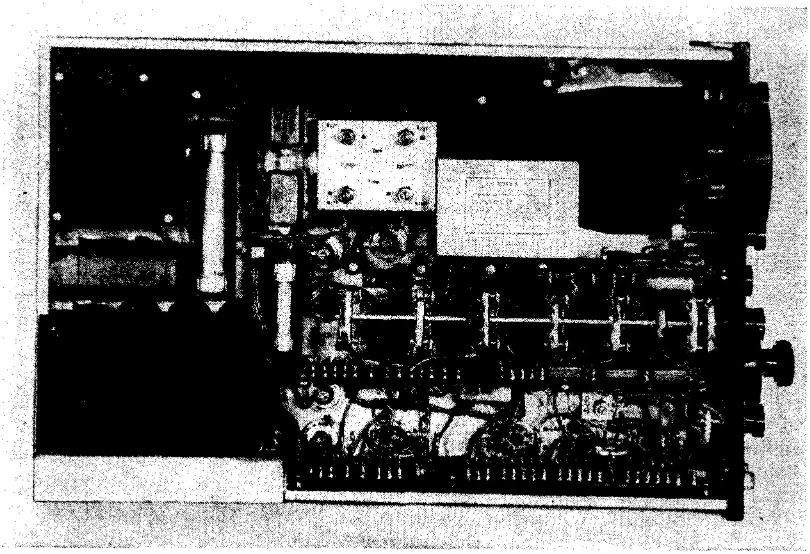


FIG. 18·24.—Left interior view of the A/R range scope.

is 1-5CP1, 2-2X2, 1-5U4G, 2-6AC7, 2-6H6, 9-6SN7, and 1-807. Panel connectors are of the UHF type.

The 256-B range scope was designed to stand a 10g shock. Consequently, the vertical chassis structure is of steel strongly reinforced with corner braces and structural framing. An effort was also made to achieve a maximum of accessibility and serviceability of all parts. Since the instrument was intended for battle service, the panel clearances around operating controls are large. This feature and the accessibility result in a unit of large size. All transformers, filter reactors, and capacitors larger than 0.01 μf in this unit are hermetically sealed. The remaining components are "moisture resistant," and the entire unit is anti-fungus treated.

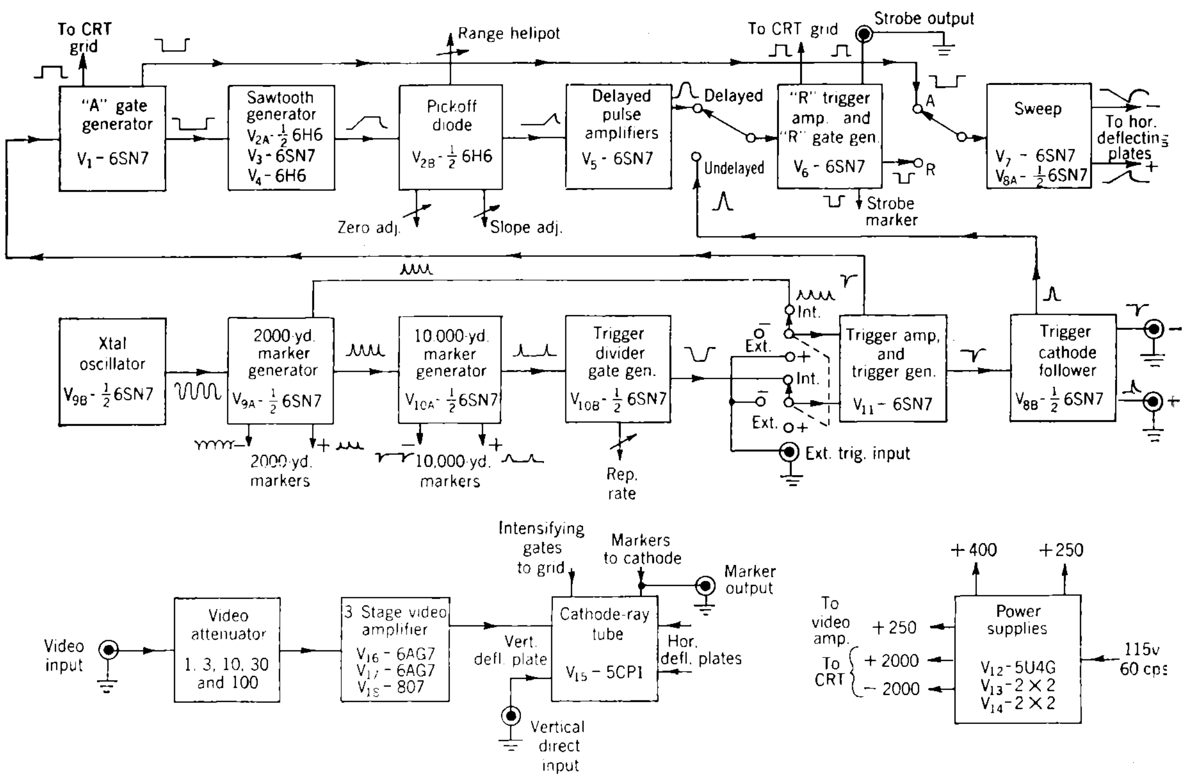


FIG. 18·25.—Block diagram of the A/R range scope.

Circuit Description.—The *cathode-ray tube circuits* in this instrument use conventional high-voltage supplies employing half-wave rectifiers followed by *RC*-filters. Focusing and beam intensity are controlled by means of potentiometers in the negative supply bleeder system. Negative bias voltage for the video amplifier is also obtained from this bleeder.

Essentially distortionless pattern centering is accomplished by symmetrically displacing each plate of the horizontal or vertical pair from a nominal voltage equal to one-half of the low supply voltage (250 volts). Dual linear potentiometers are used for this function. Another potentiometer, the auxiliary focus control, is used to adjust the second anode potential with respect to the deflecting plates. This control is located on the front panel, since it must be adjusted for each sweep rate to obtain optimum focus.

The cathode-ray tube *sweep voltages* are obtained by charging the sweep rate condensers between switch S_{1C} and ground with a feedback-type constant-current generator. Tube V_{8A} is the feedback cathode follower which supplies the positive-going CRT sweep voltage. A negative-going sweep voltage is obtained by inverting the positive sweep in a unity-gain voltage feedback amplifier, V_{7A}. Triode V_{7B} receives a negative gate either 4000, 20,000, or 200,000 yd long through switch S_{1E} which is ganged to the sweep rate selector switch S_{1C}. When this tube V_{7B} is gated "off" it allows the sweep to start and causes it to be restored when it is "on." During the time V_{7B} is off the cathode-ray tube is intensified. Consequently, on the 800- and 2000-yd sweeps the trace runs off the right-hand side of the tube.

The A sweep multivibrator, which actuates the cathode-ray tube sweep generator and the range delay sawtooth generator, is the self-biased direct-coupled single-time-constant multivibrator V_1. The negative output of this multivibrator, taken from the plate load of V_{1A}, is used to turn the linear sweep clamp tube V_{3A} off and, via switch S_{1E}, to gate V_{7B} as previously described. The positive output pulse, taken from the plate load of V_{1B}, intensifies the cathode-ray tube on the 20,000- and 200,000-yd scales.

The R sweep delay and *ranging circuit*, comprised of V_2, V_3, and V_4, uses a "bootstrap"-type constant-current generator which charges a capacitor to obtain a linear sweep. This circuit is elaborated by the use of clamping diodes to relieve the load on the bootstrapping cathode follower V_{3B} and to obtain rapid recovery of the charged condensers upon termination of the linear-sweep period. This is accomplished by diodes V_4 and, in part, V_{2A}. The sweep condensers for the 20,000- and 200,000-yd range are each split into two temperature-compensated series groups. The lower of these groups has an auxiliary charging path through the resistors across each section of V_4 to the cathode of V_{3B}. Current flowing in this circuit is approximately proportional to the square of the elapsed time since the circuit was triggered. These resistors have been chosen so that this current balances out the second-order current flowing in the charging resistor (between V_{2A} and S_{1F}) due to imperfect bootstrapping. By this compensation, a sweep linear to better than 0.1 per cent is obtained at the grid of V_{3B}.

The pick-off diode V_{2B} has the linear-sweep voltage impressed on its plate and a variable bias derived from a 10-turn precision potentiometer (which is

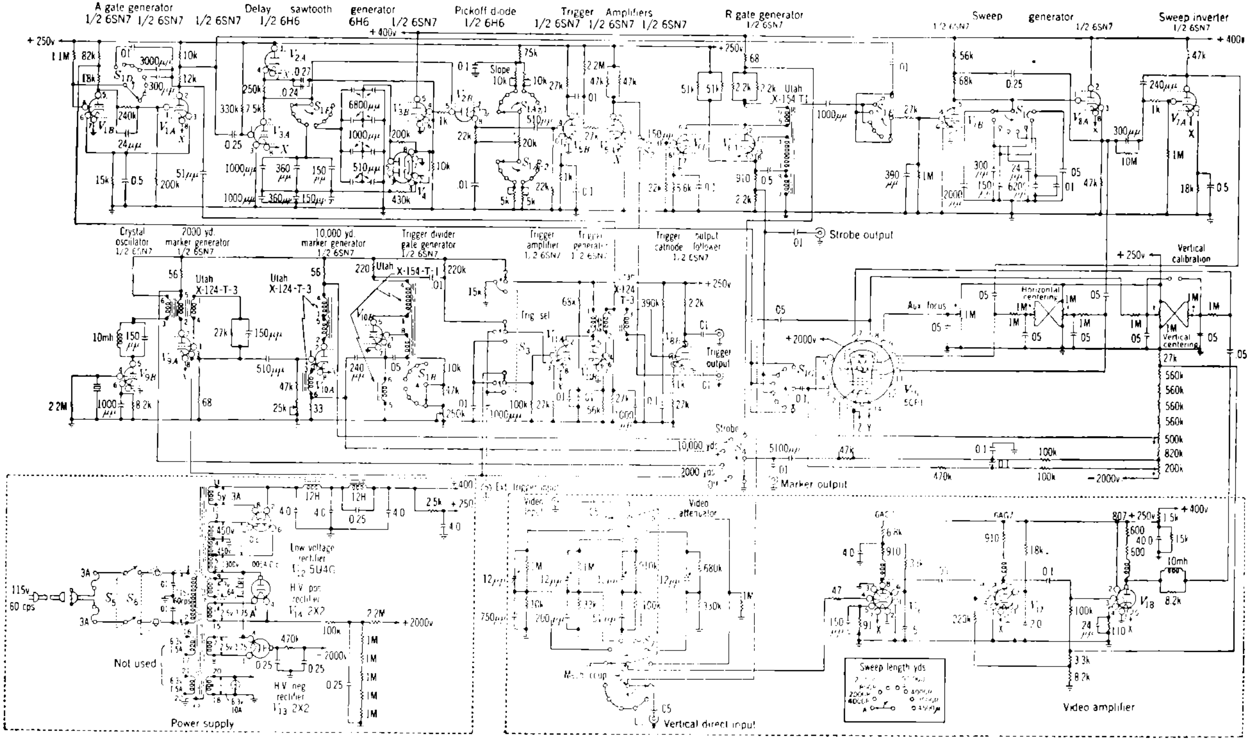


FIG. 18·26.—A/R range scope circuit.

coupled to the range dial) on its cathode. When the sweep voltage equals the cathode bias voltage, V_{2B} conducts, and a portion of the sawtooth voltage passes over to the delay amplifier V_5. Here it is amplified until its rate of rise is sufficient to give stable firing of the 4000-yd R gate blocking oscillator V_{6A}. Slope and zero potentiometers are provided for setting the calibration of the range dial on the 20,000- and 200,000-yd delays. These controls are accessible through a side panel in the case.

A switch, located on the front panel, permits firing the R gate generator directly from the trigger. This allows the R gate to cover the range between the minimum linear delay and the trigger. As previously mentioned, the negative R gate output via switch S_{1E} starts the short sweeps, and the positive output via switch S_{1G} intensifies the CRT during the sweep interval.

The *timing standard* for the range mark generator is an 81.94-kc/sec oscillator V_{9B}. The Class C plate current pulses of this tube are used to synchronize the 2000-yd marker generator blocking oscillator V_{9A}. Low-impedance short range marks are derived across the current pulse resistors in the plate and cathode circuits. A division by 5 to obtain 10,000-yd range marks is accomplished in a plate-to-cathode coupled blocking oscillator V_{10A}. This type of circuit was chosen for its stability with respect to heater voltage variations. As in the previous case, current pulse outputs are obtained across the resistors in the plate and cathode circuits of V_{10A}. The final repetition rates are obtained from a gate-generating blocking oscillator V_{10B} and are controlled by the panel-mounted potentiometer in the grid circuit of this stage. The minimum PRF of this circuit is about 80 cps while the upper rate is limited to about 2000 cps by the fixed grid circuit resistance. On the 200,000-yd delay position of S_1, deck S_{1B} introduces additional series resistance in the timing circuit, limiting the maximum PRF to 400 cps. This is done to avoid exceeding the maximum duty ratio of the linear-sweep delay circuit.

A jitter-free trigger is obtained by selecting one of the 2000-yd range marks of V_{9A} with the output pulse of V_{10B}. This is accomplished in the coincidence tube V_{11A} which is so biased (approximately twice cutoff) that no plate conduction occurs unless the negative cathode gate derived from the plate circuit of V_{10B} coincides with (that is, brackets) a positive grid pulse derived from the cathode circuit of V_{9A}. Coincidence causes V_{11A} to draw a pulse of plate current for the duration of the positive 2000-yd mark, and the normally off blocking oscillator V_{11B} fires. Its output through the paraphase amplifier, V_{8B}, furnishes positive and negative triggers at the trigger output panel connectors. The internal A gate trigger is obtained from the timing condenser of the blocking oscillator V_{11B}. This point gives an unambiguous negative trigger upon differentiation in the coupling condenser making the firing of the A gate reliable. When the oscilloscope is triggered externally, the triggers are shaped in V_{11} to obtain essentially constant triggers for the A or R gates. For external triggers, S_3 converts V_{11A} to a biased amplifier by removing the positive bleeder-bias connection. Positive input triggers are then grid-coupled with the cathode resistor bypassed. Negative triggers are introduced in the cathode circuit with the grid resistor bypassed. In each case the incoming trigger fires the blocking oscillator V_{11B}. A shaped output trigger synchronous with the input trigger can be obtained from the trigger output connectors. Switch S_3 also disables the crystal oscillator and marker generator when an external trigger is used. This prevents the possibility of improper use of the range marks with unsynchronized triggers.

The *video amplifier* is preceded by a four-step constant input impedance capacity-compensated resistance attenuator. Each attenuator section is independent. The video amplifier, with the exception of the 807 output stage, is a shunt-peaked Class A amplifier. The output stage uses both shunt and series peaking. Cathode peaking is also employed in the input and output stages.

The high-voltage *power supply* uses two half-wave rectifiers V_{13} and V_{14} to supply both positive and negative 2000 volts at 5 ma for the cathode-ray tube circuits. The low-voltage supply, which uses V_{12} as a full-wave rectifier, delivers 200 ma at a potential of 400 volts. Of this current, 60 ma is dropped to 250 volts in an *RC*-filter section. The second stage of the two-stage *LC*-filter is shunt-tuned to approximately 120 cps.

18·5. Oscilloscope TS100/AP.[1] *Function.*—Test oscilloscope TS100/AP has been designed to perform two general functions. It is a precision range calibrator which uses a circular sweep to measure the time intervals normally encountered in radar practice with an accuracy of about 0.02 per cent. It is also used with linear-type "A" sweeps as a general-purpose portable test oscilloscope for radar equipments.

Characteristics.—A type 3DP1 *cathode-ray tube* operating at an accelerating potential of 1800 volts is used. This gives a vertical deflection factor of about 150 volts/in. Two *sweep generators* are included in this instrument, providing gated or continuous circular sweeps and triggered linear sweeps. The length of each revolution of the circular sweep is 12.361 μsec (1 nautical mile), accurate to 0.02 per cent. This sweep is generated continuously. All revolutions or the first 1, 30, or 350 revolutions immediately following the synchronizing trigger may be intensity-gated. Triggered linear-sweep lengths of approximately 10, 370, and 4300 μsec (0.8, 30, and 350 nautical miles) are provided. Either the intensifying gate of the circular sweeps or any of the linear sweeps may be delayed by a continuously variable period to a maximum of 620 μsec (50 nautical miles) with respect to the synchronizing trigger.

[1] Developed by the Radiation Laboratory and manufactured by the United Cinephone Corporation, Torrington, Conn. Procured by the Signal Corps, Dayton Procurement Office; and Radiation Laboratory. The *Manuscript of Handbook of Maintenance Instructions for Test Oscilloscope TS100/AP* is an instruction manual printed by the manufacturer. Also available is the technical order TO-08-TS100/AP-2, published by the Signal Corps, Dayton Procurement Office. The sections on maintenance and the waveforms shown in this latter publication are inaccurate and do not apply to this instrument. This unit is used with the AN/APS-15 and other airborne radar systems. The Model 5 type AJ test oscilloscope (60ACZ) which is described briefly at the end of the present section is of similar mechanical and electrical design.

The sweep circuits may be synchronized either by the internal trigger or by an externally generated trigger of either polarity having an amplitude of 15 to 50 volts and a minimum rate of rise of 100 volts/μsec.

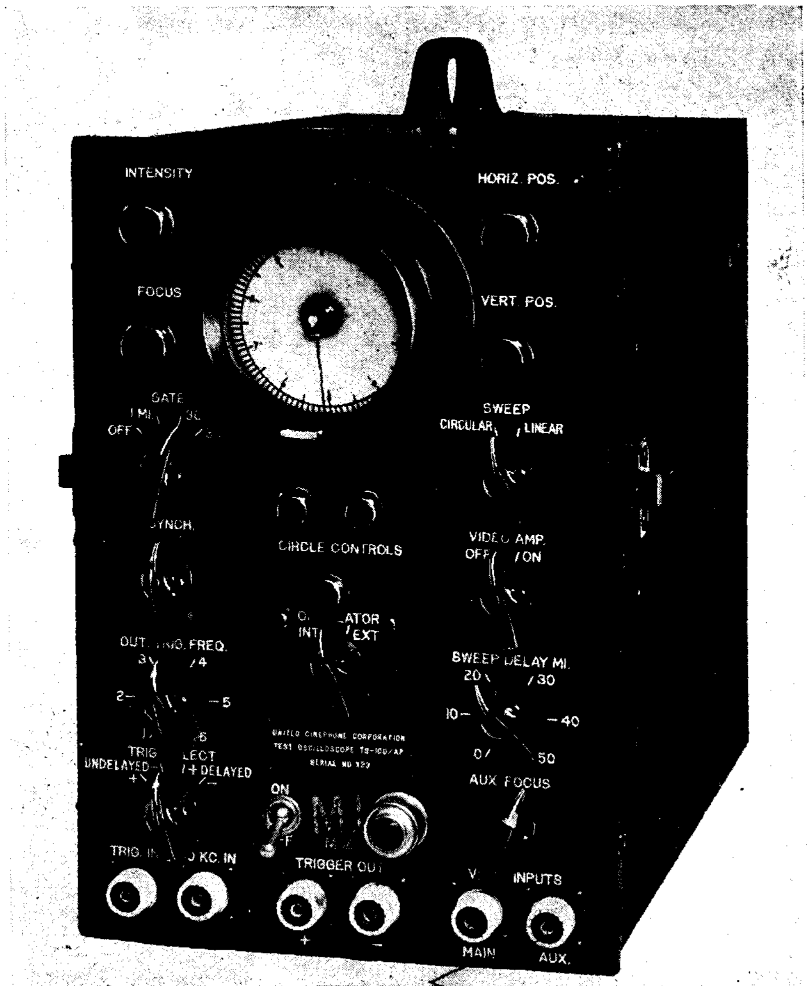


FIG. 18·27.—Oscilloscope TS100/AP.

Satisfactory triggering may be obtained up to 5000 cps when the sweeps are triggered directly or to 1200 cps when the variable delay is used. A positive linear-sweep voltage of 225 volts amplitude may be obtained from a connector on the back of the instrument. This circuit has an internal impedance of 100,000 ohms. An input is also provided on the panel for an externally generated sinusoidal voltage having an amplitude of 35 to 100 volts rms and a frequency between 78 and 84 kc/sec to generate a circular sweep.

The circular sweep, which is generated by quadrature voltages from a crystal oscillator, provides a *precision time base* for the measurement of short time intervals. The period of each cycle, which corresponds to the time for one revolution of the sweep, is accurate to better than 0.02 per cent. The position of a signal on the circle can normally be read to $\pm.05$ μsec (± 8 yd). Marker pips generated by the crystal oscillator can also be made available by adding a panel connector and one wire to the instrument.

A 100-volt positive trigger and a 70-volt negative trigger having a pulse width of 0.6 μsec are supplied by the *trigger generator* for synchronizing external equipment. These triggers are accurately locked in phase with the circular sweep by frequency division. The repetition rate is variable from 300 to 1500 EM/SEC in integral submultiples of 80.86 kc/sec. A low-impedance output is provided.

A direct *signal channel* to either or both vertical deflecting plates is provided when the linear sweep is used or to the center electrode for the circular sweep. The input impedance is 1 megohm paralleled by 55 μμf for the upper vertical deflecting plate and 45 μμf for the center electrode. The *video amplifier* has a voltage gain of about 12 which is constant to ± 3 db between 250 cps and 2.8 Mc/sec. Overloading occurs with an input signal amplitude greater than 2 volts. The input impedance is 1 megohm paralleled by 32 μμf, and the stage gain is fixed with no control provided. Only one signal input is provided for the circular sweep. Signals can be mixed on the linear sweeps, however, as inputs to both vertical deflecting plates are provided. One input is connected directly and the other either connected directly or through the video amplifier.

The over-all dimensions of this oscilloscope are 9 by 14 by $16\frac{1}{2}$ in., and the weight including the cover is 42 lb. Test cables weighing a total of 2 lb are also supplied. Power requirements are 115 $\pm$ 10 volts (the transformer may be reconnected for operation on 230 volts), 50 to 1200 cps, 110 watts at 60 cps. Ambient temperature limits are $-22°$ to $+122°$F. The tube complement is 1-3DP1, 1-2X2, 1-5Y3GT/G, 1-6AG7, and 7-6SN7GT. Space is provided in a removable panel cover for the variety of test leads and cables supplied.

Circuit Description.—The *cathode-ray tube control circuits* use a conventional high-voltage bleeder containing the focus and intensity potentiometers. Dual potentiometers provide balanced centering voltages for deflecting the beam. An

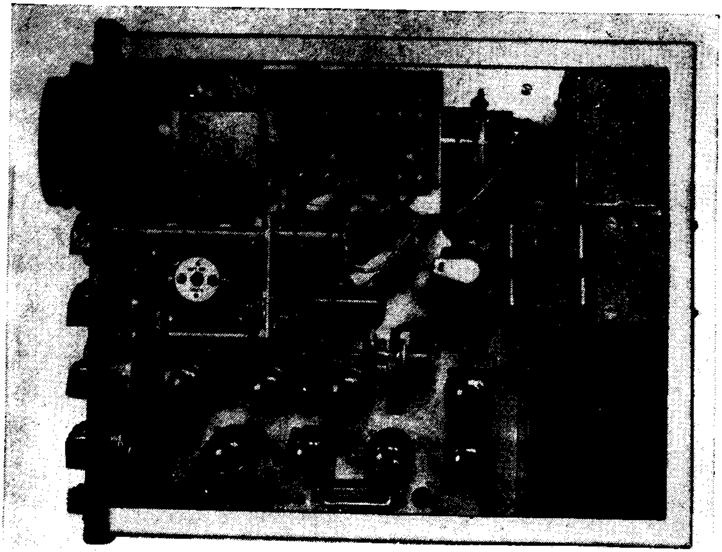


FIG. 18·28.—Right interior view of the TS100/AP.

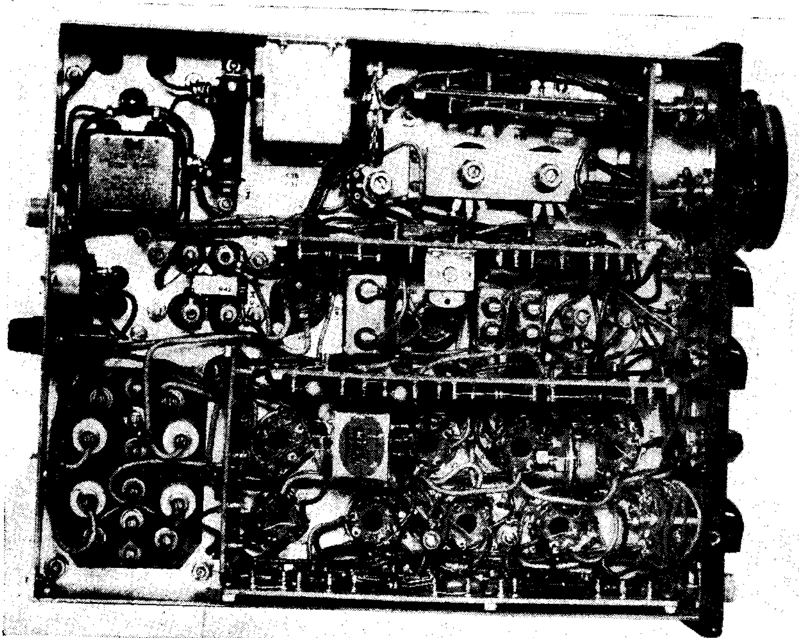


FIG. 18·29.—Left interior view of the TS100/AP.

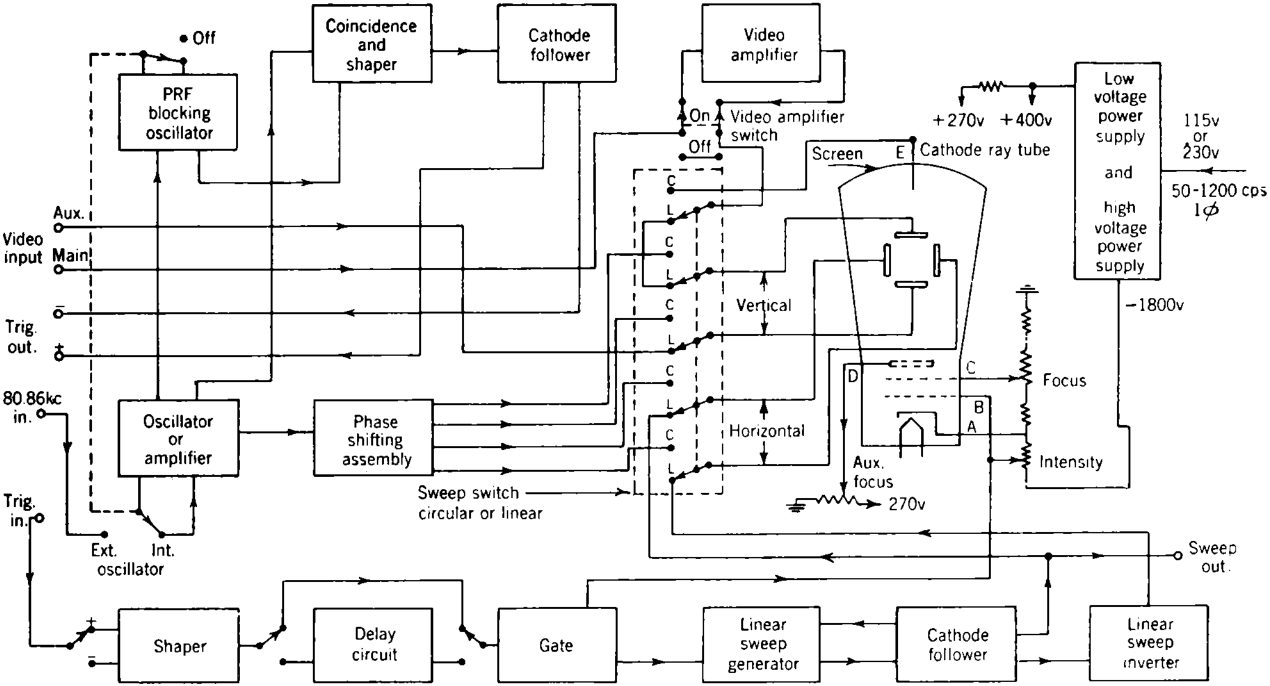


FIG. 18·30.—Block diagram of the TS100/AP.

astigmatism control, which sets the voltage on the center electrode and the second anode, is mounted on the panel, since it must normally be adjusted to obtain good focus when changing from a circular to a linear display.

The *circular sweep* is generated by a triode crystal oscillator V_{7B}. The crystal is ground to a frequency of 80.86 kc/sec giving a period that is equivalent to one radar nautical mile. The oscillator may also be used with an 81.94-kc/sec crystal for a 2000-yd period. Several circuit changes are necessary to modify this instrument for use with a 100-kc/sec crystal (giving a 10-μsec period).

The tuned transformer in the oscillator plate circuit (Sickles RE10001) has two sets of secondary windings which are tuned by panel-mounted air trimmer condensers paralleled by temperature-compensating fixed capacitors. These windings are each coupled to a pair of cathode-ray tube deflecting plates and are tuned on opposite sides of the resonant frequency to give a 90° phase difference between the voltages across them. When the horizontal and vertical deflection amplitudes are also equal, a circular trace is obtained. The circle diameter is controlled by a trimmer condenser across the primary. It may also be adjusted by the screw-driver-controlled potentiometer in the oscillator plate circuit to compensate for variation in the deflection sensitivity when the cathode-ray tube is replaced.

When the oscillator switch S_2 is in the EXT. position, the crystal is shorted out and the oscillator stage functions as a low-gain amplifier. A sinusoidal voltage from an external oscillator having either of the frequencies mentioned above may then be connected to the 80-kc IN connector to generate the circular sweep.

The circuits generating the *linear sweep* include a trigger shaper, a sweep delay multivibrator, a sweep gate multivibrator, a "bootstrap" sweep generator, and an inverter to provide push-pull deflection. The trigger shaper consists of the triode tube sections V_{1A} and V_{1B}. The first tube receives either a positive trigger on the grid or a negative trigger on the cathode and amplifies it to fire the biased-off blocking oscillator V_{1B}. The waveform of the pulse produced by this stage is essentially independent of the input trigger characteristics and provides the optimum trigger for the multivibrator following.

The two triode sections of V_2 are used in a cathode-coupled delay multivibrator whose pulse length is set by the SWEEP DELAY grid-voltage control. When the TRIG SELECT switch S_1 is in the + or − DELAYED position, this multivibrator is triggered by the blocking oscillator pulse. The fall of the positive rectangular pulse that is generated at the plate of V_{2B} is differentiated to provide a delayed trigger for the sweep gate multivibrator. With switch S_1 in the + or − UNDELAYED position, the delay multivibrator is not used and the sweep gate multivibrator is triggered directly by the trigger shaper. The + and − positions of the switch S_1 refer to the input trigger polarity. The SWEEP DELAY control has a roughly calibrated dial to indicate the time delay in terms of nautical miles. Slope and zero adjustments can be made with the screw-driver-adjusted potentiometers in the control voltage bleeder to make the actual delay agree with the dial reading. This delay is normally used with the 1-mile circular or linear sweep to expand a portion of the viewed signal.

Tube sections V_{5B} and V_{5A} make up the sweep gate multivibrator which is

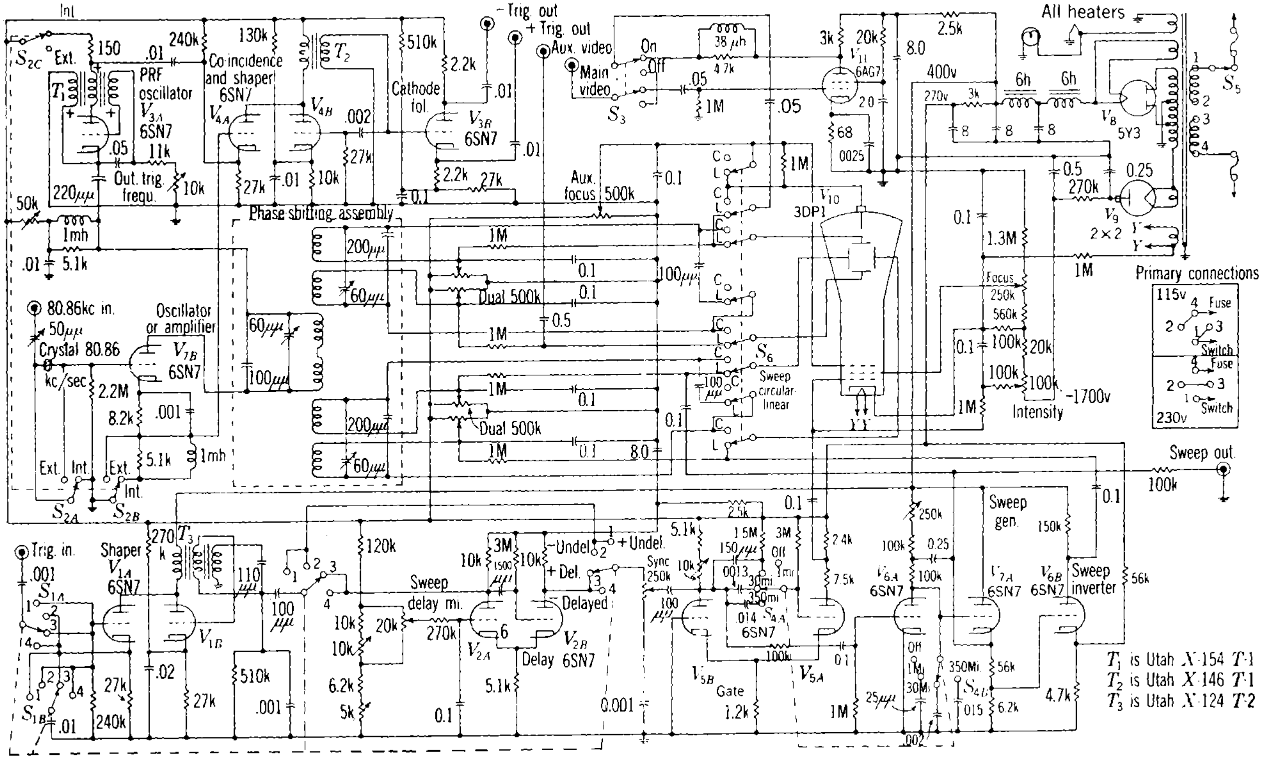


FIG. 18·31.—TS100/AP circuit.

also of the cathode-coupled type. The length of the pulse produced by this circuit is controlled by changing the feedback capacity with the GATE switch S_{4A}. The positive pulse developed at the tap in the plate-load resistor of V_{5A} is coupled to the grid of the cathode-ray tube to turn on the beam when the sweep voltage is generated. The negative pulse from the plate of V_{5B} is coupled to the grid of the clamp tube V_{6A}. Simultaneous control of the three pulse lengths selected by switch S_{4A} is provided by the screw-driver-adjusted potentiometer in the plate circuit of V_{5B}. This makes it possible to compensate for the variation of pulse length with tube and component changes. The SYNC potentiometer is used to adjust the trigger amplitude when the time interval between trigger pulses is less than the multivibrator gate length. Stable operation at these high repetition rates may be obtained when the trigger amplitude is just sufficient to cause the multivibrator to fire.

The linear-sweep voltage is generated by the clamp tube V_{6A} and the cathode follower V_{7A}, which are used in a bootstrap-type circuit. Since the sweep duration is set by the length of the gating pulse, the capacity of the sweep condenser must be changed for each gating pulse length if the sweep amplitude is to be constant. The GATE switch S_{4B} performs this function. Simultaneous adjustment of all sweep amplitudes can be made with the screw-driver-adjusted potentiometer in the plate circuit of V_{6A} to compensate for tube and component variation. The positive sweep voltage is coupled to one horizontal deflecting plate from the cathode follower. An attenuated voltage of similar waveform is developed at the tap in the cathode-load resistor of this stage and is coupled to the grid of the inverter amplifier V_{6B}. The negative sweep voltage at the plate of this tube is approximately equal in amplitude to the positive sweep voltage and is coupled to the opposite horizontal deflecting plate.

A single-stage *video amplifier* is connected into the main video signal input circuit when the VIDEO AMP switch S_3 is in the ON position. Series peaking and cathode peaking are used to extend the bandwidth. No gain control or attenuator is necessary, since the gain is quite low ($G = 12$). The amplifier is connected to either the top vertical deflecting plate or the radial deflecting center electrode by the SWEEP switch S_6, which selects either linear or circular sweep operation. Direct connection to both vertical deflecting plates or to the center electrode may be made through the MAIN VIDEO and AUX VIDEO inputs when the video amplifier switch S_3 is in the OFF position.

The *trigger generator* is designed to provide positive and negative output triggers which are accurately locked in phase with the sinusoidal voltage generating the circular sweep. Any phase shift, or "jitter," in this relationship would cause a broadening of the signal trace or an actual shift in its position on the circular sweep. Since time measurements are made by direct comparison with a scale on the face of the cathode-ray tube, the accuracy of the instrument would be materially reduced.

By inserting small resistance-damped inductances in the plate and cathode circuits of the crystal oscillator, relatively short voltage pulses (or pips) are produced by the pulses of plate current in the Class C oscillator. The pips produced in the plate circuit are used to synchronize an otherwise free-running blocking oscillator V_{3A}. This circuit oscillates at a rate determined by the discharge time constant of its grid circuit which is varied by the OUT. TRIG. FREQ. potentiometer. The circuit constants of this blocking oscillator have been so chosen that a relatively square negative current pulse of about 15 volts amplitude and 18-μsec duration is generated across the resistor in the plate circuit. This pulse and the positive 15-volt pips generated in the cathode circuit of the crystal oscillator are coupled to the cathode and grid respectively of the coincidence amplifier V_{4A}. The fixed bias level of this tube is set to give conduction when these pulses occur simultaneously. Since the crystal generated pips have a 12-μsec period, this will occur for the pip immediately following the one that causes the blocking oscillator to fire. When coincidence occurs, the amplified pip appears as a negative trigger at the plate. Since V_{4A} has a common plate circuit with the blocking oscillator V_{4B}, this stage is fired. The 0.6-μsec positive

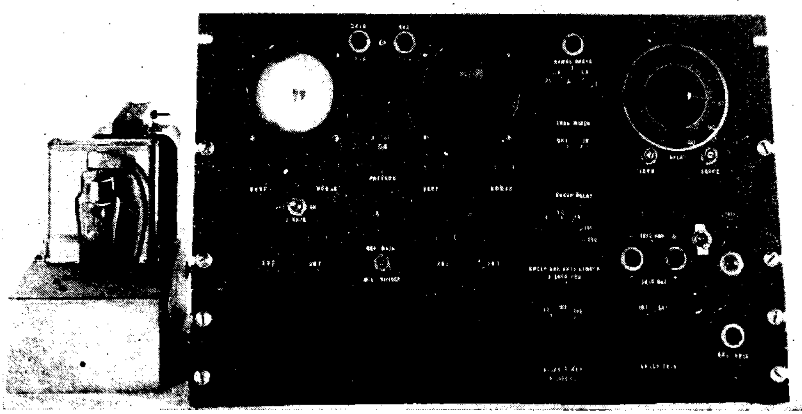


FIG. 18·32.—Laboratory type A and J oscilloscope.

trigger generated is coupled to the cathode follower inverter V_{3B} which supplies low-impedance positive and negative output triggers.

A *power supply* of conventional design is used. The low-voltage supply gives 400 volts at 40 ma and 270 volts at 45 ma. A −1700-volt, 2-ma supply is provided for the cathode-ray tube. The power transformer has a two-section primary winding which may be connected for operation on either 115 or 230 volts.

Similar Equipments.—The Model 5 type AJ test oscilloscope manufactured by the Technical Apparatus Company of Boston, Mass., is almost identical with the TS100/AP in electrical and mechanical design. The circuits have been modified, however, to provide a 6.1-μsec (1000-yd) circular sweep and linear-sweep lengths of approximately 5-, 370- and 2500-μsec duration. The repetition rate range of the trigger generator has been changed to 250 to 1000 cps. Other characteristics of the two instruments are essentially the same.

Only component value changes were necessary in the oscillator, sweep gate multivibrator, and sweep generating circuits. The Sickles transformer No. RE13386 was developed to supply the quadrature voltages for the circular sweep.

A redesign of the blocking oscillator frequency divider V_{3A} was necessary to generate the rectangular 8-μsec current pulse required for selection of one of the pips generated in the 163.88-kc/sec crystal oscillator. The same pulse transformer is used, however. No changes were required in the other circuits.

While the redesign was made for a specific application, this unit has some advantages over the TS100/AP. The most important of these are higher reading

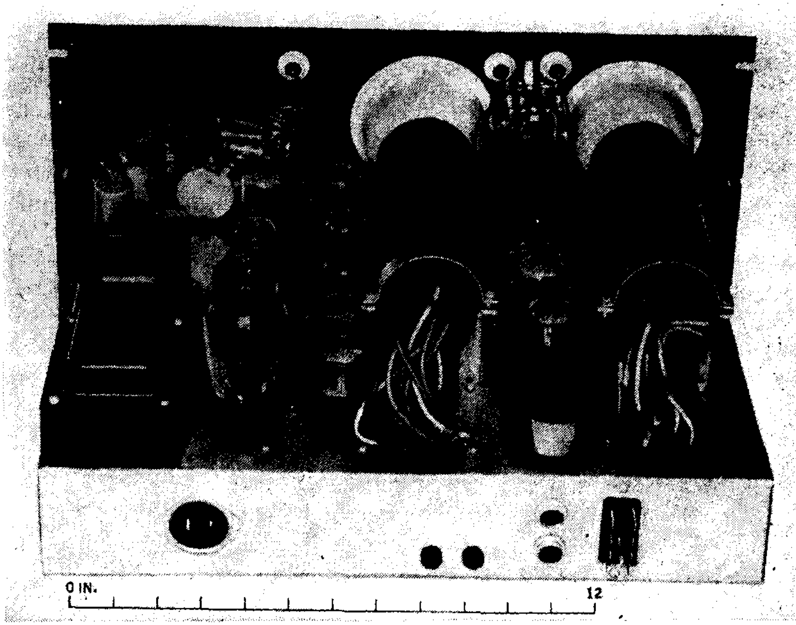


FIG. 18·33.—Top view of the laboratory model.

accuracy obtained with the 1000-yd circular sweep and the faster linear sweep for viewing portions of waveforms.

Figures 18·32 and 18·33 show a laboratory instrument that used the same circuits as the TS100/AP. Somewhat greater flexibility was obtained by using separate cathode-ray tubes for the linear and circular sweeps, eliminating switch S_6. Signal inputs go to both tubes simultaneously, removing some of the ambiguity encountered in the use of the circular sweep alone.

18·6. Model III Range Calibrator.[1] *Function.*—The Model III range calibrator is a relatively specialized device for accurately measuring small time intervals in the laboratory. It uses a circular sweep generated by

[1] Developed by the Radiation Laboratory and the F. W. Sickles Co. and manufactured by the latter company. Procured by the Radiation Laboratory and others. "Operating Data for the Model III Calibrator" is an instruction report published by the F. W. Sickles Co. The Models I and II Sickles Calibrators (RL Report No. 333) are earlier models of this unit which have similar functions.

a crystal oscillator as a precision time base and supplies synchronizing triggers to external signal generating equipment. Most of its functions are duplicated by the circuits associated with the circular sweep in the TS100/AP test oscilloscope (Sec. 18.5).

Characteristics.—A type 3DP1 *cathode-ray tube* is used. This tube has a signal electrode in the middle of the screen for producing radial deflection of a circular time base. An accelerating potential of 2000 volts is employed, giving a radial deflection sensitivity of about 225 volts per inch for a positive signal on a 2.5-in. diameter circular sweep.

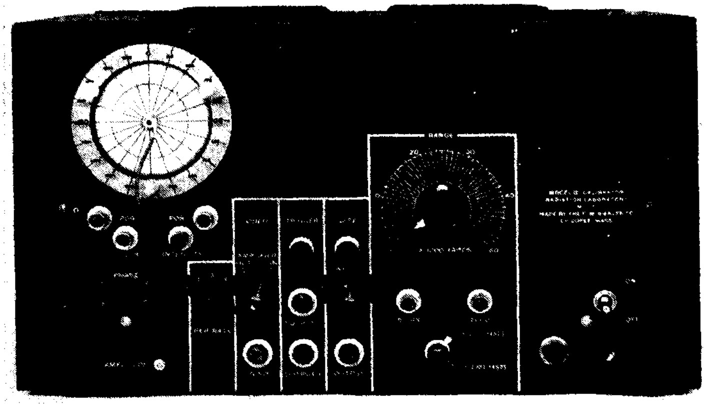


FIG. 18·34.—Model III range calibrator.

Either continuous (type J) or gated (type X) circular sweeps having a fixed frequency are provided by this instrument. Three models having sweep periods of 10, 10.74 (1 radar statute mile), and 12.2 μsec (2000 yd) per revolution have been constructed. The corresponding oscillator frequencies of 100, 93.11, and 81.94 kc/sec are generated with an accuracy of about 0.02 per cent. Pulse position on any given sweep revolution can be read to about ± 0.1 μsec with the circular scale on the CRT screen. The circular sweep generating circuits cannot be triggered but must supply a synchronizing trigger to external equipment. A single revolution of the circular sweep may be intensified and delayed with respect to the output trigger. When properly calibrated, the delay should agree with a calibrated range dial to within ± 1 per cent. Higher accuracy is not necessary, since it is necessary to have only an indication of the particular sweep revolution being illuminated. The maximum delays for the three models are 300 μsec, 25 land miles, and 50,000 yd. A delayed output trigger may be derived from the narrow (intensifying) gate output if desired.

The internal *trigger generator* provides positive and negative output triggers having an amplitude of 100 volts and a pulse length of 0.6 μsec. These pulses are locked in phase with the crystal oscillator and are developed across resistances of 1000 ohms. Nominal repetition rates of 400, 800, 1200, 2000, and 2300 pps are selected by a five-position switch.

A sinusoidal output voltage from the oscillator that generates the circular sweep may be obtained from a connector on the rear of the chassis.

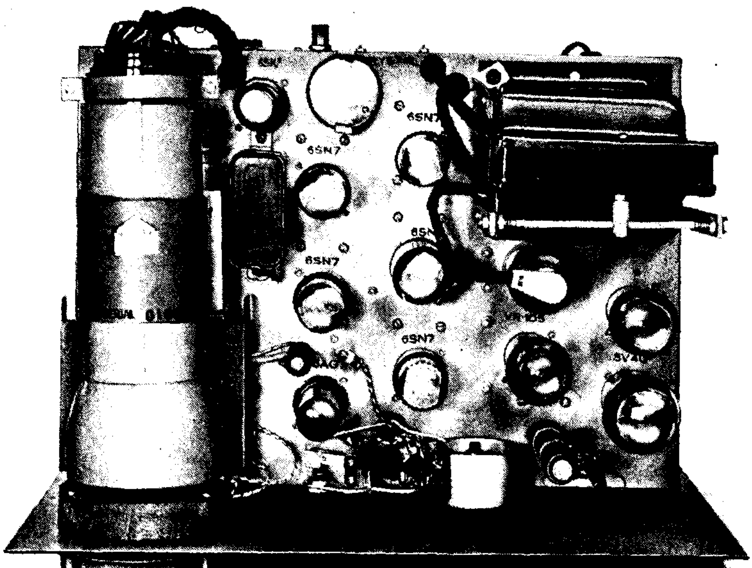


FIG. 18·35.—Chassis view of the Model III range calibrator.

While no input connector is provided, an externally generated sinusoidal voltage may be used to generate a circular sweep. The oscillator, which should have a frequency near that for which the unit was designed, should be connected through a 25-μμf condenser to the internal oscillator grid pin of the socket from which the crystal has been removed.

The *signal channel* to the center electrode of the CRT either is a direct connection or passes through a video amplifier. The input impedance with the amplifier switched out is 750,000 ohms paralleled by 37 μμf. The single-stage *video amplifier* has a voltage gain of approximately 18 which is constant within ±3 db between 50 cps and 4 Mc/sec. Overloading occurs with an input signal exceeding 3.5 volts peak. The amplifier input impedance is 750,000 ohms paralleled by 25 μμf.

The over-all dimensions of this unit are 11 by 17 by 10 in. high, and it weighs 40 lb. Power requirements are 117 ± 8 volts, 60 cps, 110 watts. The tube complement is 1-3DP1, 1-2X2/879, 1-5V4G, 1-6AG7,

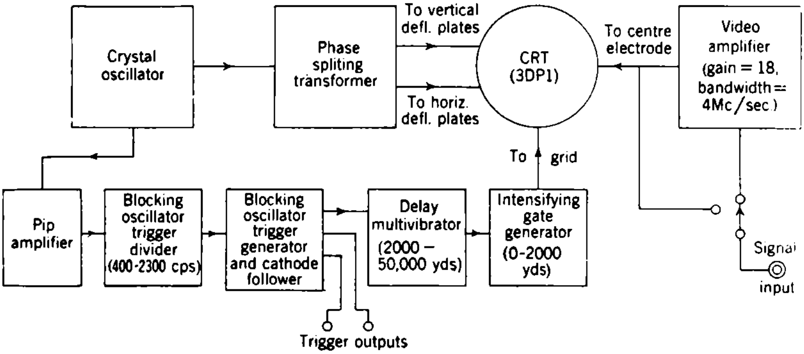


FIG. 18·36.—Block diagram of the Model III range calibrator.

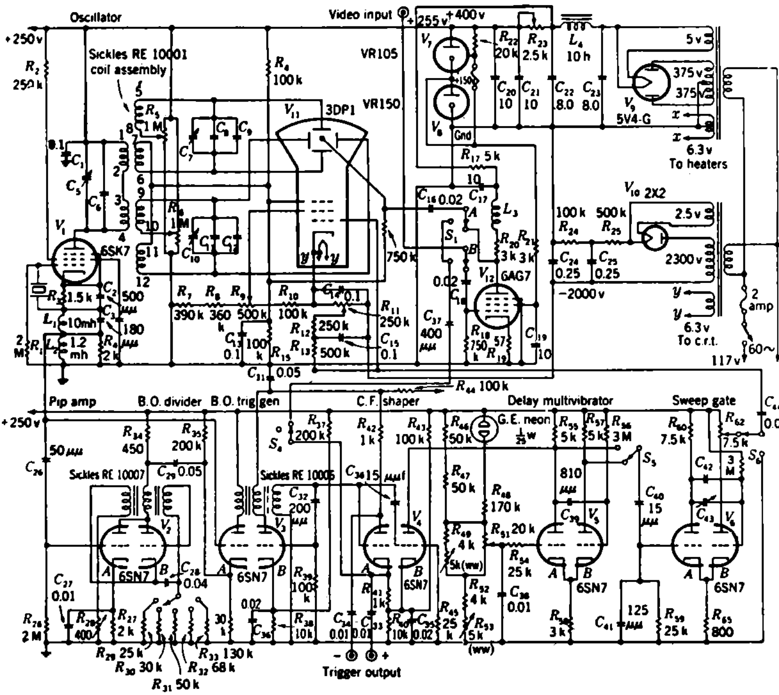


FIG. 18·37.—Model III range calibrator circuit.

1-6SK7, 5-6SN7, 1-VR105, and 1-VR150. The only cable provided is a nonremovable 6-ft power cord. Panel connectors are Amphenol type 80C.

Circuit Description.—The *cathode-ray tube circuits* include a high-voltage bleeder circuit of conventional design which contains the FOCUS and INTENSITY potentiometers. Centering voltages are applied to one of each pair of deflecting plates through the RE10001 coil assembly. Either a direct-coupled or an amplified signal may be applied to the center electrode via the VIDEO switch S_1. Positive intensifying pulses are coupled to the grid of the CRT.

Tube V_1 is used in a crystal oscillator *sweep generating circuit* which produces a voltage output of stable frequency across the primary of the tuned transformer in the plate circuit. The secondary tuning condensers may be adjusted to give sinusoidal output voltages having a 90° phase difference across the two sets of secondary windings. Each of these voltages is coupled to a pair

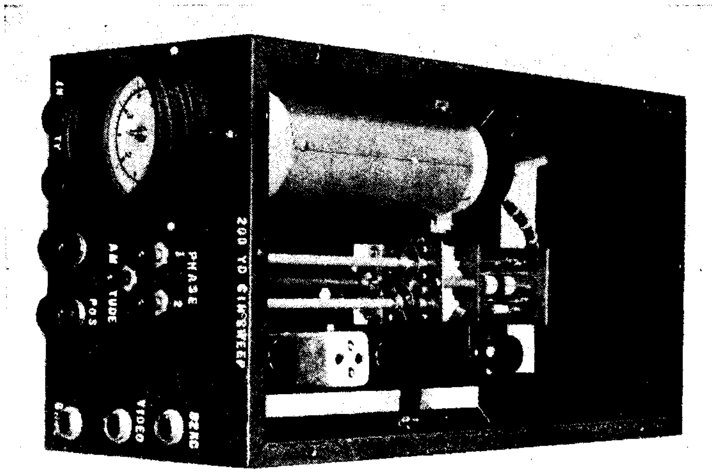


FIG. 18·38.—200 yard "J" scope attachment.

of cathode-ray tube deflecting plates, generating a circular trace with a diameter controlled by the primary tuning condenser.

The measurement of a longer time interval than that required for one revolution of the circular sweep requires an auxiliary circuit to indicate the number of revolutions made by the sweep before the signal is observed. This is accomplished by using a calibrated delay circuit to delay a pulse whose duration is somewhat less than the period of the sweep. This pulse is applied to the grid of the cathode-ray tube. In this manner, any particular revolution of the sweep, out to the maximum delay of 25 to 30 revolutions following the trigger, can be intensified. The range of the illuminated revolution is indicated on the calibrated delay control.

The circuits used include a trigger shaper, a delay multivibrator, and a sweep gate multivibrator. The positive trigger from the blocking oscillator is coupled to the grid of the trigger shaper V_{4B}. The output of this stage fires the delay multivibrator V_5. The length of the pulse produced by this stage may be adjusted with the RANGE control which varies the bias voltage on the grid of the normally off tube. Slope and zero controls may be adjusted to make the pulse length agree with the control calibration. The neon tube in the control bleeder reduces the effect of changes of plate voltage on the gate length.

Either the positive rise at the start of the pulse at the first plate or the positive fall of the end of the pulse at the second plate may be coupled through switch S_5 and an *RC* differentiating circuit to the first grid of the sweep gate multivibrator V_6. This stage produces a positive pulse, the length of which is adjusted by the GATE WIDTH condenser. It is coupled through the GATE switch S_6 and a high-voltage coupling condenser to the grid of the cathode-ray tube. The first revolution of the sweep after the trigger is intensified when switch S_5 is in the 0-2000

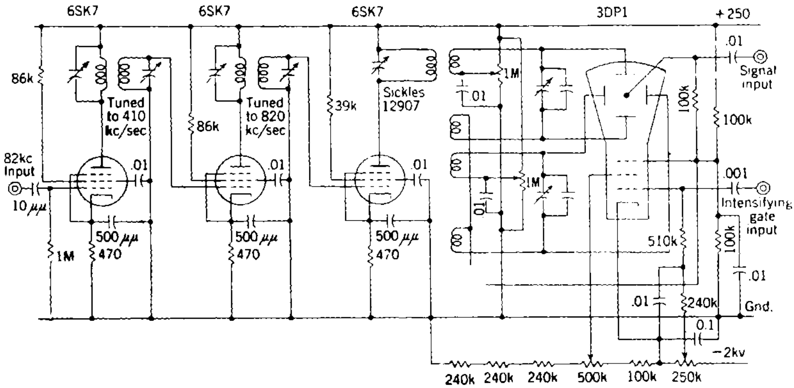


FIG. 18·39.—Circuit of the 200-yd "J" scope.

YARDS (or equivalent) position. In the 2000 YARDS position the intensifying gate delay is controlled by the RANGE knob.

The internal *trigger generator* provides positive and negative output triggers which are accurately locked in phase with the sinusoidal voltage generating the circular sweep.

Tube V_7 is used in a single-stage low-gain *video amplifier* circuit. Signals may be introduced into the video amplifier input circuit or bypassed around the amplifier by the VIDEO AMPLIFIER switch S_1. Shunt peaking in the plate circuit is used to extend the bandwidth. No gain control is provided.

When a higher reading accuracy than that obtained with the 80- to 100-kc/sec sweep is required, an auxiliary 800- to 1000-kc/sec sweep may be used. A design that has seen considerable use with the Model III calibrator is shown schematically in Fig. 18·39. It consists of two amplifier stages whose plate circuits are tuned to give a total frequency multiplication of 10, a high-frequency phase-splitting transformer (Sickles No. 12907), and a cathode-ray tube with its associated controls.

18·7. The TS126/AP Test Oscilloscope.[1] *Function.*—The TS-126/AP test oscilloscope is a small portable synchroscope and pulse generator which was designed for field maintenance and calibration of the ranging circuits of certain airborne radar systems. A simple cable adapter permits its use as a general-purpose instrument. It is useful

FIG. 18·40.—Oscilloscope TS126/AP.

for signal tracing and the measurement of short time intervals in pulse circuits.

Characteristics.—The *cathode-ray tube* is a type 2AP1 which operates at an accelerating potential of 1200 volts giving a vertical deflection factor of 200 volts per inch. The tube face is recessed about 1 in. behind the panel, and a green light filter is provided.

The triggered *sweep generator* uses a high-impedance audio choke for linearization which is center-tapped to give push-pull sweep voltages.

[1] Developed by the Radiation Laboratory and the Galvin Mfg. Corp. and manufactured by the latter company. Procured by the Signal Corps Dayton Procurement Office as part No. 54P3964. The "Preliminary Instruction Book for Range Calibrator TS-126/AP" was published by the Galvin Mfg. Corp. This unit is used with the AN/APG-5, 8, 14, and 15 radar systems.

Sweep lengths of 25- and 125-μsec duration are provided. This unit is designed to operate with trigger recurrence rate between 950 and 1650 cps but may be modified for rates outside this range. Either the internal trigger generator or an external trigger of 40 volts or greater amplitude may be used to synchronize the sweep generator.

Positive marker pips having a spacing of 2.44 μsec (400 yd) are generated by a gated oscillator in this unit and serve as a *timing standard.* They may be coupled internally to the input of the video amplifier or to external equipment from one pin of a cable fitting on the oscilloscope panel. The time interval between the trigger and the first marker is adjustable from 3.4 to 5.2 μsec (550 to 850 yd). A 409.50-kc/sec crystal filter is used to adjust the oscillator frequency accurately. This frequency is determined by a slug tuned *LC* resonant circuit and may be changed readily if desired.

A negative trigger having an amplitude of 100 $\pm$ 10 volts may be supplied to external equipment by the internal *trigger generator.* The output pulse has a rise time of 0.18/μsec, a duration of 1/μsec, and a repetition rate variable from 950 to 1450 cps. It is designed to operate into a minimum load impedance of 4000 ohms paralleled by 250/$\mu\mu$f.

A positive pip of 10 volts maximum amplitude which may be delayed by a variable amount out to 49 μsec (8000 yd) is also provided. It may be coupled internally to the video amplifier and to external equipment having an input impedance of about 1000 ohms. The purpose of this marker is to check the ability of the system under test to lock on and track a moving target. With the addition of a blocking oscillator and cathode follower, it can be used to supply a delayable output trigger. The delay control knob is not calibrated, but the delay is linear to about 1 per cent with its rotation.

A *signal channel* to the top vertical deflecting plate is connected either directly or through one of two high-voltage capacity dividers. The input impedance for direct connection is 1 megohm paralleled by 40 $\mu\mu$f. Deflection factors of 200, 1000 and 4000 volts per inch are available. The *video amplifier* has a gain of 25 and a bandwidth between 0.707 gain points of 2.3 Mc/sec. Overloading occurs with a signal exceeding $\pm$3 volts peak amplitude. The input impedance is 147,000 ohms paralleled by 25 $\mu\mu$f. These values are obtained with the attenuator switch on position 5. A five-step low-voltage attenuator is provided for the video amplifier, giving vertical deflection factors of 370, 100, 30, 15, and 8 volts per in. Since the video amplifier is coupled to the lower deflecting plate, two channels are available for mixing signals.

This oscilloscope is $10\frac{1}{4} \times 10\frac{1}{2} \times 14\frac{1}{4}$ in. over all including the front cover, and the weight without cables is 32 lb. Power requirements are

115 ± 10 volts, 400 to 1600 cps, 82 watts at 400 cps. Ambient temperature limits are −67° to 122°F. The tube complement is 1-2AP1, 1-2X2, 1-5Y3GT, 2-6AC7, 1-6SL7, and 5-6SN7. For use as a general-purpose oscilloscope, one of the large cables should be cut and connected to the

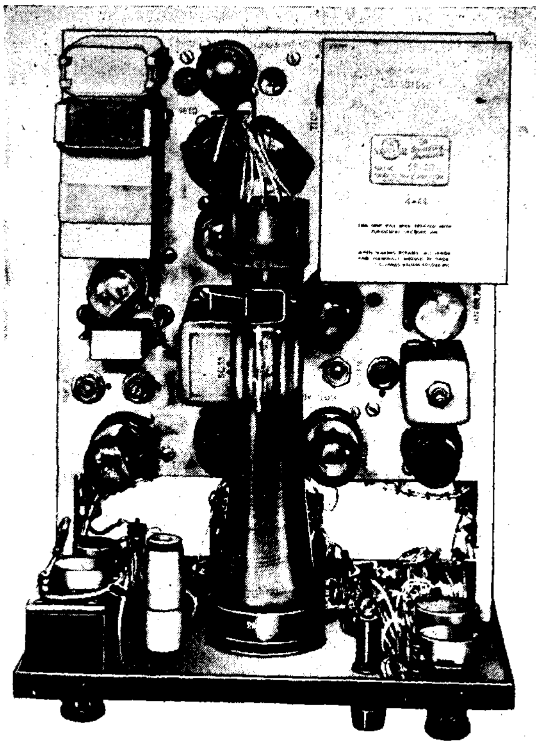


FIG. 18·41.—Chassis view of the TS126/AP.

junction box shown in Fig. 18·44. A panel cover with space for the small test leads and space fuses is included.

Circuit Description.—The *cathode-ray tube circuits* employ a conventional high-voltage bleeder circuit which includes the focus and intensity potentiometers. Dual centering potentiometers provide balanced d-c voltages for the deflecting plates. Push-pull sweep voltages are used; but since the signal deflections are normally small, single-ended inputs suffice for the vertical deflecting plates.

The *sweep circuits* include a multivibrator, a clamp tube, and a large choke for linearizing the charging rate of the sweep condenser. The multivibrator employs tube sections V_{1A} and V_{2B} in a circuit that has a quiescent state and produces only one cycle of oscillation for each trigger pulse supplied. These triggers may be obtained from the internal trigger generator or from an external source. Rectangular voltage pulses having a duration determined by the setting of the sweep switch S_{1B} are generated at each plate. This switch controls the pulse duration by changing the plate-to-opposite-grid coupling capacity.

The negative pulse from the plate of V_{2B} is coupled to the grid of the clamp tube, causing this tube, which has been conducting a relatively large current, to be biased off. The sweep condenser, whose size is determined by the sweep switch

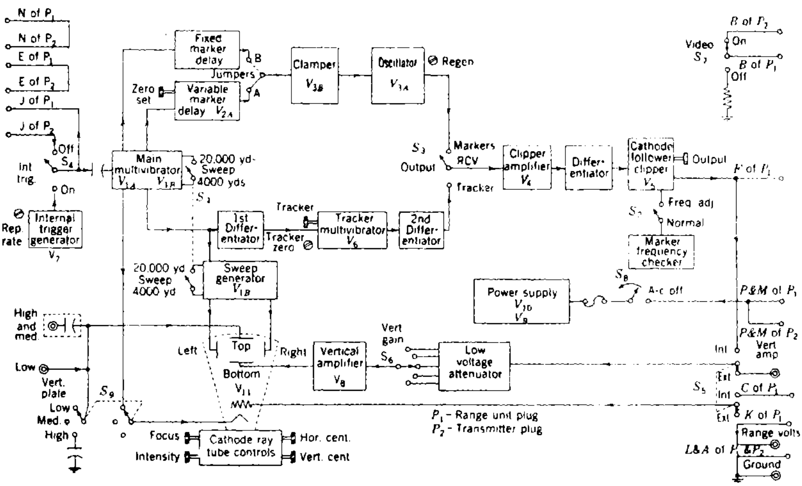


FIG. 18·42.—Block diagram of the TS126/AP.

S_{1A}, then charges through the inductance in the plate circuit of V_{1B} generating a positive linearly rising pulse. A voltage of approximately equal amplitude but of opposite polarity is induced at the opposite end of this center-tapped inductance. These push-pull voltages are coupled to the horizontal deflecting plates of the cathode-ray tube giving balanced deflection of the beam. When the vertical plate switch S_{9A} is in the LOW position, the beam is intensified by the attenuated multivibrator pulse generated at the tap in the plate resistor of V_{2B}.

A shocked oscillator is used as a *timing standard* in this unit. With the delay circuit jumpers in the "B" position, the negative multivibrator pulse from the plate of V_{2B} is coupled to the grid of the clamp tube V_{3B}. Cutting off the current flowing in this tube starts oscillation of the parallel resonant circuit in its cathode circuit. The amplitude of oscillation is maintained at a constant level by positive feedback from the cathode follower V_{3A} for the duration of the multivibrator pulse. A slug-tuned inductance is used to adjust the oscillator frequency, while the screw-driver-adjusted potentiometer in the cathode circuit of V_{3A} controls the amount of feedback. The sinusoidal oscillator output is fed through switch S_{3A} to the clipper-amplifier V_4. The output of this stage is differentiated by the inductance in the grid circuit of the output cathode follower V_5. This cathode follower has a positive quiescent cathode potential set by the OUTPUT poten-

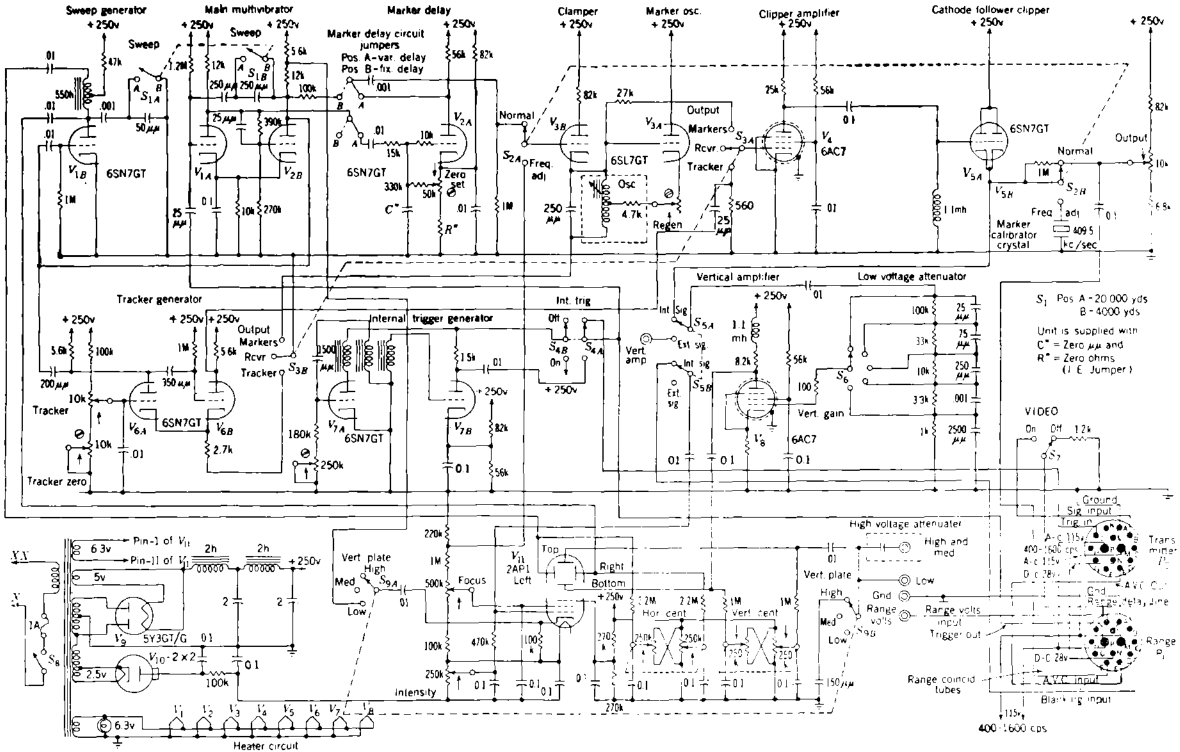


FIG. 18·43.—TS126/AP circuit.

tiometer. All portions of the input signal having a lower potential level are clipped, giving an output of a train of positive marker pips whose amplitude is determined by the setting of the potentiometer. These may be coupled to the video amplifier or to the output connector.

When the marker delay circuit jumpers are in the "A" position, the positive multivibrator pulse is fed to the grid of the marker delay tube V_{2A}. The pulse rise time is reduced materially by the resistance-capacity network in the grid circuit. The grid-to-cathode potential may be varied below cutoff by adjusting the ZERO SET potentiometer so that a definite period of time elapses before the grid potential rises above cutoff and conduction begins. This delays the output pulse, which is coupled to the oscillator clamp tube, and the marker pips by a like amount.

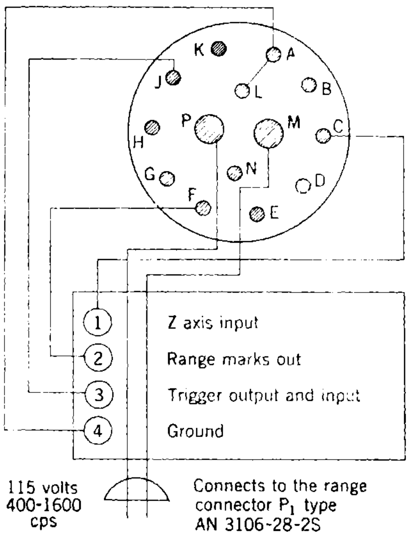


FIG. 18·44.—Cable adapter.

When switch S_2 is thrown from the NORMAL to the FREQ. ADJ. position, the oscillator bias is changed to permit continuous oscillation. Also, a filter crystal having a maximum impedance at 409.5 kc/sec is connected in the cathode circuit of V_5. When the oscillator is tuned to this frequency, a large increase in output voltage is observed, providing an accurate check of the oscillator frequency.

A delayed marker generator is also included in the design of this oscilloscope. The negative multivibrator pulse is differentiated by a short time constant RC-circuit, and the resulting negative pulse is used to trigger the delay multi-vibrator V_6. The duration of the pulse produced by this circuit is an essentially linear function of the grid voltage of V_{6A}. This potential is controlled by the TRACKER potentiometer which is mounted on the panel. The screw-driver-adjusted TRACKER ZERO potentiometer provides an adjustment of the minimum pulse length. The positive pulse produced by this circuit is differentiated, amplified, peaked, and clipped in V_4 and V_5 as described previously for the timing oscillator. A single positive pip of variable amplitude is obtained from the cathode-follower output and is coupled to the video amplifier and output connector. This pip, which is delayed with respect to the system trigger, may be used directly as a video signal or may trigger an external blocking oscillator and cathode follower circuit to generate a delayed trigger.

An internal *trigger generator* uses V_{7A} as a free-running blocking oscillator and V_{7B} as clipper amplifier to give a fairly low impedance negative output trigger. When the INT. TRIG. switch is in the ON position, plate voltage is supplied to V_7

and the trigger is coupled to both the main multivibrator and the output connector. A screw-driver-adjusted variable resistance in the grid circuit of the blocking oscillator is used to set the repetition rate.

A single type 6AC7 pentode V_8 is used as a low-gain *signal amplifier.* The input signal may be introduced from an external source through a connector on the panel or from the internal marker generators through switch S_{5A}. A five-position resistance-capacity attenuator is included for gain control. Shunt peaking is used to extend the video bandwidth. The amplifier output is capacitively coupled to the lower vertical deflecting plate of the cathode-ray tube.

A 400-cycle power supply of conventional design is used to supply voltages of 250 volts of 105 ma and -1100 volts at 2 ma. Transformer and filter redesign are necessary for 60-cycle operation.

18·8. Direct-coupled Oscilloscope for Potentiometer Testing.[1] *Function.*—This oscilloscope[2] was originally designed to provide a visual indication of the departure from linearity of precision wire-wound potentiometers. To do this, the potentiometer under test is mechanically coupled to a master potentiometer of considerably higher accuracy and both are connected across a 10-volt d-c supply. The voltage at the arm of the master potentiometer deflects the CRT beam horizontally by an amount proportional to the shaft rotation, while the difference voltage between the arms of the potentiometers is amplified to produce the vertical deflection. By calibrating a scale on the CRT screen in terms of per cent nonlinearity, the potentiometer characteristic for all shaft positions is obtained directly.

A later circuit design considerably extended the usefulness of the instrument by including a simple thyratron sweep generator. This made it generally applicable to the wide variety of applications requiring an oscilloscope with high-gain direct-coupled amplifiers.

Characteristics.—A type 5CP7 *cathode-ray tube* (having a long persistence screen) is operated at a total accelerating potential of 4000 volts. The latter voltage is obtained from an audio oscillator supply to provide voltage regulation, stabilizing the deflection sensitivity of the cathode-ray tube. The sweep voltage may be obtained from the internal thyratron *sweep generator*, or an externally generated sweep voltage having an amplitude of at least 10 volts may be used. Sweep rates of approximately 2 to 10,000 cps are provided in five ranges.

The gain of the *direct-coupled differential amplifier* is adjustable in steps covering a range of linearities of 0.1 to 5 per cent of full scale (1-in. deflection on the CRT screen). The corresponding differential input

[1] Section 18·8 is by B. Chance and H. J. Reed, Jr.

[2] Developed and constructed at the Radiation Laboratory; RL drawings Nos. A-14295A and A-14602A are the original schematics for this oscilloscope.

potentials for potentiometers supplied with 10 volts are 10 to 500 mv. The bandwidth of 10 Mc/sec is adequate for following the rapid fluctuations of linearity normally encountered. Less than 1 mv unbalance in the output voltage results from variation of the input potential from 0 to 10 volts. The gain of the amplifier is sufficiently stable that the output scale may be calibrated directly in percentage accuracy and

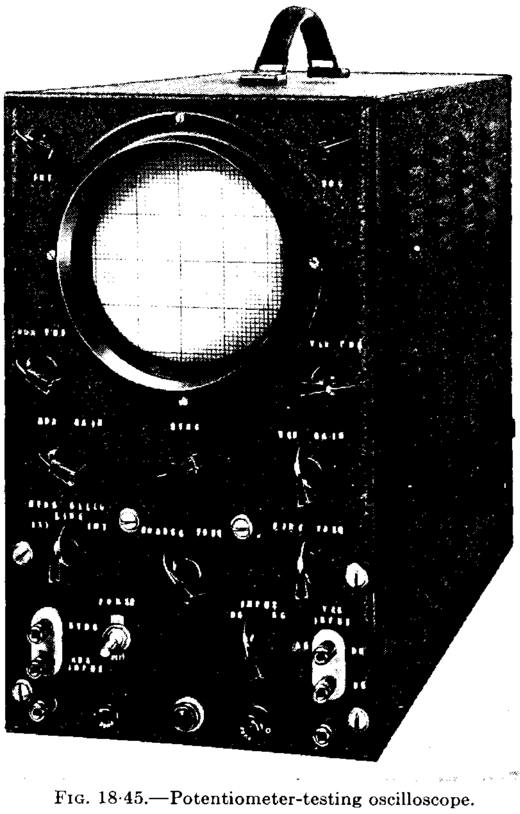


FIG. 18·45.—Potentiometer-testing oscilloscope.

readings to 0.01 per cent linearity may be made. Drift is small enough that measurements can be made over a period of several hours without readjusting the zero.

The over-all dimensions of this oscilloscope are 8 by 13½ by 19 in., and the weight is 50 lb. Power requirements are 115 ± 10 volts, 60 cps, 100 watts. Tubes used are 1-5CP7, 1-5Y3, 1-6B4, 1-6SH7, 4-6SL7, 2-6SU7, 1-6V6, and 1-6X5.

Circuit Description.—The *cathode-ray tube circuits* include a conventional high-voltage bleeder, containing the focus and intensity controls, and an auxiliary focus potentiometer which sets the second anode potential. Direct coupling to all deflecting plates is employed so the horizontal and vertical centering controls adjust the bias levels in the sweep and signal amplifiers.

A type 884 thyratron, V_6, is used in a relaxation oscillator to generate a positive-going sawtooth *sweep voltage.* Coarse frequency control is obtained by switching the sweep condensers with S_2, while a 4-megohm potentiometer in the charging circuit affords fine frequency control. The sawtooth voltage is coupled through the cathode follower V_{7A} to the cathode coupled amplifier inverter V_8. Positive and negative sawtooth voltages are obtained at the plates of this stage

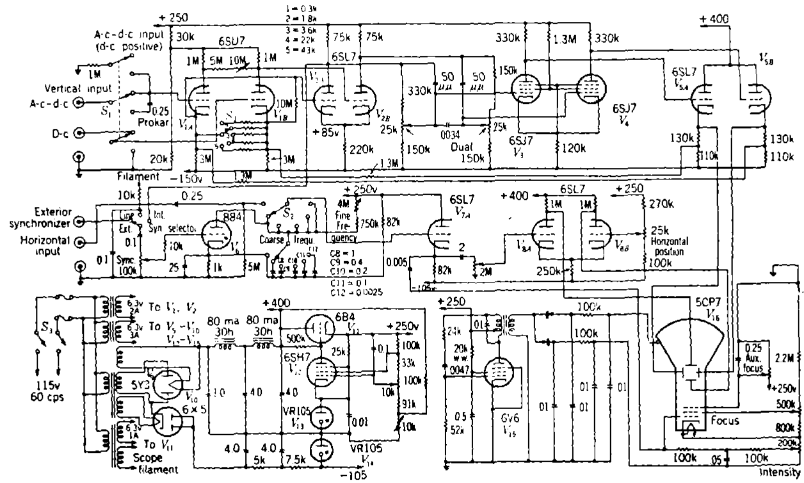


FIG. 18·46.—Circuit of potentiometer tester.

and are coupled directly to the horizontal deflecting plates of the CRT. The potentiometer in the grid circuit of V_{8A} is a sweep amplitude control, while the one in the grid circuit of V_{8B} controls the horizontal centering. The intensification pulse for the CRT is obtained by *RC* differentiation of the cathode-follower output. An external sweep input to the grid of the cathode follower is provided on the first position of switch S_2.

A three-stage direct-coupled *differential amplifier* is used to drive the vertical deflecting plates of the cathode-ray tube with over-all negative feedback used to obtain the high order of stability required. The first stage is, of course, the most critical. The cathode-coupled differential amplifier used has its plate voltage reduced by a bleeder, permitting the use of comparable values of cathode and plate resistors. With the values chosen the plate current is low, roughly 100μa. The cathode resistor is split into two parts having resistances of 3 megohms each. The balance of the tube sections V_{1A} and V_{1B}, pretested to 0.5 per cent for type

6547, is further improved by the 10-megohm balance control between the plates which is adjusted for optimum rejection of "common-mode" grid-bias changes.

The plates of V_{1A} and V_{1B} are directly coupled to the grids of V_{2A} and V_{2B}, and the plates of the latter are coupled to the grids of V_3 and V_4 through potential dividers. A dual 25-k potentiometer is inserted in this network to compensate for small differences in the characteristics of V_{1A} and V_{1B}. This control is used as a means of adjusting the trace of the oscilloscope to the center of the screen when the input is shorted and should not be used as a positioning control. Tubes V_3 and V_4 are pentodes to provide sufficient gain and to give a large output voltage. Cathode followers V_{5A} and V_{5B} are interposed between pentodes V_3 and V_4 and the cathode-ray tube deflecting plates to avoid capacitance loading and to provide a low-impedance driving point for the feedback attenuator.

Negative feedback is established over the three amplifier stages by connection from the output cathode follower to the appropriate cathodes of the first stage. If these cathodes are short-circuited, there is no feedback for differential signals. The over-all gain may be adjusted by varying the resistance between these cathodes, and resistors associated with switch S_1 have been selected to give gains corresponding to between 0.1 to 5 per cent nonlinearity for a full-scale deflection on the CRT. Direct feedback to the cathodes is preferred to voltage addition in the grid circuits, as both input grids are completely unencumbered.

The result of supply voltage variation on the amplifier balance is indicated in Table 18·1. In all cases except the variation of the unregulated supply, the

TABLE 18·1. EFFECT OF VOLTAGE VARIATION ON ZERO POINT OF TRACE ON OSCILLOSCOPE. DEVIATIONS ARE IN MILLIVOLTS (MEASURED WITH GAIN = 15,000).

	Unregulated supply		Regulated supply			Negative supply			Heater voltage		
Supply voltage	250	400	225	250	275	95	105	115	5.7	6.3	6.9
Amplifier output	0	0	+5	0	−4	−3	0	+3	+2.5	0	−1.0

fluctuations have been impressed upon the system artificially. Stabilization of the positive and negative supply voltages and the heater voltage by methods to be discussed reduces the fluctuations due to external line voltage variation to a negligible amount. The remaining limitations on the stability of the d-c amplifier are initial drift which lasts approximately 30 sec and long-time drift due to differences in the aging of V_{1A} and V_{1B}. With high-impedance input signal sources, the practical limitation in the sensitivity of the oscilloscope is set by electrostatic pickup from the high-voltage oscillator transformer, and a few millivolts of ripple are seen on the oscilloscope display at maximum gain with 10-megohm input impedances. This effect is negligible with inputs from the usual potentiometers.

The gain-frequency characteristic of the amplifier without feedback and with 15-$\mu\mu$f phase-advance condensers between the plates of V_2 and the grids of V_3 and V_4 is indicated in Fig. 18·47. With feedback, however, the bandwidth is cut down to approximately 10 kc by a condenser between the control grids of V_3 and V_4 in order to avoid oscillation with high feedback ratios.

The low-voltage *power supply* is conventional and employs series regulator tube V_{11}. Almost complete compensation for line voltage variation is obtained by including a portion of the bleeder for the unregulated supply in the grid circuit of the control tube V_{12}. The common resistance is a 10-k variable resistor which is adjusted to give a minimum variation of the regulated voltage. Because filament voltage variation has a large effect upon the zero point of the first two tubes, a Sola regulator transformer has been used to supply the heaters of V_1 and V_2. Operating at low currents, these regulators have given somewhat better stabilization than the quoted value (1 per cent), and with selected type 6SU7's excellent independence of line voltage has been obtained. The variation of deflection sensitivity of the cathode-ray tube with line voltage was found to be

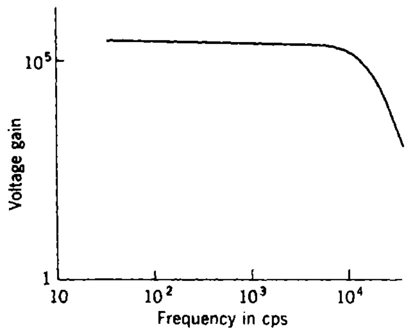


FIG. 18·47.—Gain-frequency characteristic of amplifier without feedback.

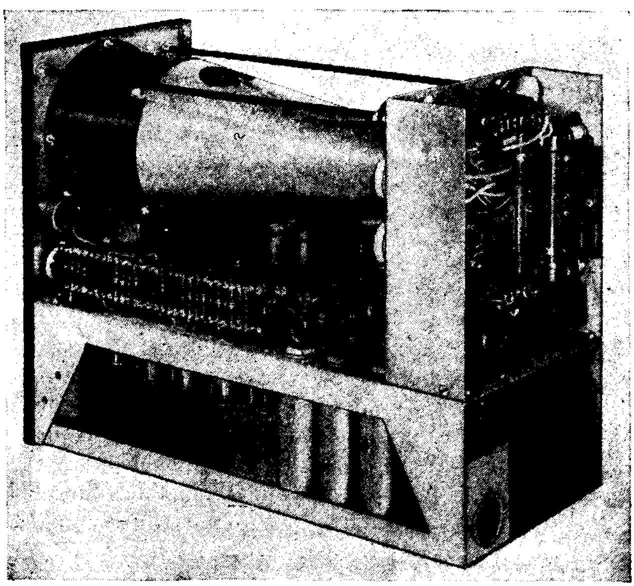


FIG. 18·48.—Top view of the potentiometer-testing oscilloscope chassis.

too great to permit the use of an unregulated high-voltage supply, so a 1-kc/sec audio oscillator V_{15} and selenium rectifiers are employed to obtain a more stable voltage. The d-c output voltages obtained are $\pm$ 2kv.

Layout details are shown in Fig. 18·48. Particular care has been given to the location and shielding of the components in the first two stages of the direct-coupled amplifier.

18·9. The TRE General-purpose Monitor.[1,2] *Function.*—This is a very useful general-purpose oscilloscope having a variety of triggered and free-running sweeps and a video amplifier that contains an electronic switch for the alternate display of two sets of information. Accurately delayed expanded sweeps are available; and since the sweep speeds and video amplifiers are calibrated, both amplitude and time measurements may be made by displacing the trace on the CRT.

Characteristics.—As design specifications for this oscilloscope were not available at the time of writing, it has been necessary to estimate them. Obtaining accurate characteristics by circuit analysis alone is impossible for many of the circuits used because of the relatively large effect of the stray wiring capacities and tube capacities. An estimate of these has been made, however, and the results should at least be of the right order of magnitude.

A type VCR-97 cathode-ray tube is operated at an accelerating potential of 3000 volts. Both triggered and free-running linear sweeps are generated by a "sanatron" circuit (Vol. 20) and a paraphase inverter to give push-pull deflection of the beam. The estimated lengths in microseconds of the triggered sweeps for the various settings of the SWEEP RANGE switch S_4 are as follows: position (1) 5, (2) 20, (3) 100, (4) 500, (5) 2500, (6) 10,000, and (7) 50,000. Free-running sawtooth sweep rates of 20 to 200,000 cps should be obtainable. An amplifier is used in the input trigger circuit, so satisfactory synchronization should be obtained with a relatively small trigger, perhaps 5 to 15 volts in amplitude. An ordinary phantastron may be switched into the trigger input circuit to provide an accurately adjustable delay in the start of the triggered sweeps. For the constants used, a maximum delay of 1500 to 2000 μsec would be expected.

Two separate video amplifiers and an electronic switch are used in the *signal channels* of this oscilloscope. Each channel consists of an input cathode follower, a three-step attenuator, and a low-gain wide-band amplifier. Assuming that the gain without feedback is very high compared with the gain with feedback, over-all gains of 5, 1, and 0.2 may be expected for the three attenuator positions. Corresponding output impedances of 1000, 200, and 40 ohms would be expected. The bandwidth should be about 6 Mc/sec for a gain of 5 and something greater than 10 Mc/sec on the other positions.

[1] Section 18.9 is by E. F. MacNichol, Jr., and H. J. Reed, Jr.

[2] Developed by the Telecommunication Research Establishment (England) and described in their specification DCDWT 1000. Circuits included with this section are TRE Diagrams No. FTR 82/971 and FTR 82/973.

Circuit Description.—The indicator circuits are shown in Fig. 18·49. Synchronization or trigger signals are applied to the inverter-cathode follower V_1. A synchronization potential of continuously variable magnitude and reversible phase is obtained from a potentiometer between the plate and cathode of V_1. This potential is applied in negative sense, positive excursion being limited by V_{2A} to the suppressor of the free-running sweep tube V_4 whose operation will be described later. When using triggered sweeps, the synchronization output at the cathode of V_1 is applied to the grid of the trigger amplifier V_3. For positive triggers, V_3 is biased off and the grid driven positive until current is limited by the 12-k grid resistor. For negative triggers V_2 is biased positively, normally drawing grid current until cut off by the trigger. The diode V_{2B} prevents positive overshoots of negative triggers from producing pulses in the plate circuit. The differentiating transformer T_1 is switched so that its output is negative for either trigger polarity. This trigger is applied either directly to the time base generator or to the sweep delay circuit.

The *delay circuit* is a conventional phantastron[1] using a VR116, V_{12}, which is triggered on the anode-clamp diode V_{10}. The screen waveform is differentiated in the transformer T_2 whose output triggers the time base. A 220 $\mu\mu$f condenser across the cathode resistor delays the change in cathode voltage, producing a more rapid change in screen current than would otherwise be obtainable.

The *sweep voltage generator* is a screen-coupled "sanatron." The Miller integrator V_7 is cut off in the quiescent condition as its screen potential is held negative by the trigger tube V_6. A type VS-70 neon lamp is used to obtain a fixed voltage drop between the plate of V_6 and the screen of V_7 regardless of the currents drawn. Triggers are injected into the grid of V_6 through the diode V_{5A} cutting it off momentarily. The plate of V_6 and screen of V_7 then rise, causing space current to flow in V_7 and starting the Miller rundown. The negative-going sawtooth voltage charges the condenser associated with switch S_{4A} and holds the grid of V_6 at -10 volts, limited by the diode V_{5B}. The rate of descent of the plate voltage is governed by the time constant associated with switches S_{4C} and S_{4D} and the voltage to which the grid resistor is returned. Precision components determine this time constant, and the grid-resistor voltage is measured by a calibrating meter to which all measurements in the instrument are referred. When triggered sweeps are used, this voltage is present giving accurately calibrated sweeps. When the free-running sweeps are used, the voltage is adjusted by the "manual velocity" control. The sawtooth pulse is paraphased by V_9, and the resultant push-pull sweep is applied through blocking condensers and d-c restoring diodes V_{15} and V_{16} to the horizontal deflecting plates of the CRT.

The plate waveform of V_6 is a positive rectangular pulse that has a large amplitude and lasts for the duration of the sweep. A small part of this waveform is used to intensify the beam of the CRT, being applied to its grid through a blocking condenser. A d-c restorer V_{17} sets the grid waveform at a fixed value so that the intensity control does not need to be adjusted when the duty ratio is changed.

The free-running time base is produced by the addition of an extra tube to the

[1] See Vol. 20 of this series.

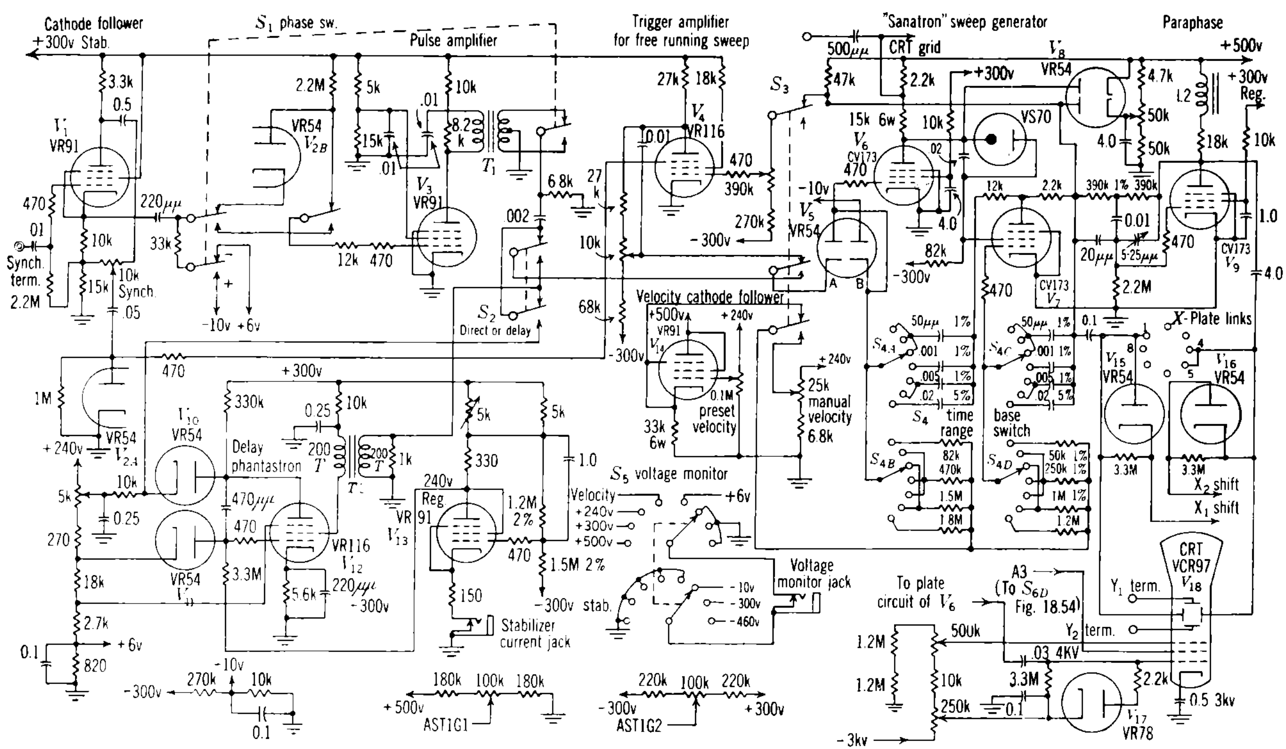


FIG. 18·49.—General-purpose monitor indicator circuit.

sanatron. In effect, this tube supplies a trigger to the regular sanatron circuit at the end of the sweep flyback. This is accomplished by applying the plate potential of V_7 through a bleeder to the grid of V_4. During the rundown V_4 is cut off as the bleeder is returned to -300 volts. This holds the cathode of V_{5A} positive with respect to its plate. Near the end of the flyback V_4 starts to conduct, dropping the cathode potential of V_{5A} until V_6 is cut off. Cutting off V_6 restarts the cycle. The negative synchronization pulses from V_1 hold V_4 off for their duration, delaying the start of the sweep until they have terminated.

The *video amplifier*, whose circuit is shown in Fig. 18·50, has two inputs which are alternately switched by a scale-of-two circuit. Signals are applied to the input attenuators either directly or through the cathode followers V_1 and V_7. The attenuators comprise the resistance-capacity dividers associated with the switches S_9 and S_{10}. As the gains of the CV-9 video amplifiers V_3 and V_4 are very high, the over-all gains are determined almost entirely by the feedback ratios. Switching is accomplished by a square wave from the scale-of-two circuit which is applied to the grids of V_3 and V_4 through the level-setting diodes V_2 and V_8 so that amplifiers V_3 and V_4 conduct on alternate half cycles. On negative half cycles V_{2A} brings the grid of V_3 negative beyond cutoff while V_{8B} is conducting. This action permits normal bias control of V_4, and the video signal on its grid is amplified. As long as the grid is not driven positive by the video input, V_{8A} is disconnected from the video circuit. On positive half cycles the situation is reversed; V_3 amplifies the video signal while V_4 is cut off. Tube V_6 is a paraphase inverter to supply push-pull deflection.

The scale-of-two circuit is essentially a multivibrator with no stable state but with coupling time constants, much longer than any expected repetition periods. To secure most rapid switching action, coupling is also achieved by means of a short time constant low-impedance feedback path to the grids which reduces the effect of the capacities to ground of the large coupling condensers. Negative triggers for the scale-of-two circuit are obtained from an external source through the diode V_{5B} which prevents positive overshoots from also operating the circuit. The triggers could be obtained by differentiating the waveforms at the plate of the gate tube of the sanatron V_6 so that the scale-of-two would transfer at the start of the flyback of the sweep. The switching transient would then be invisible as it would coincide with the time the CRT is blanked.

One vertical deflecting plate of the CRT may be connected directly to the plate of V_3, to the cathode of V_1, or to the input connector through the function switch S_6. The other vertical deflecting plate (Y_2) goes either to the plate of V_4 or to an auxiliary centering control and meter jack for signal amplitude measurements. Direct-current restoration potentials for the horizontal deflecting plates of the CRT are obtained from symmetrical centering potentiometers via leads X_1 and X_2. The potentials applied to the centering controls are switched by S_{6E} and S_{6F} so that the relationships between the potentials of the sets of deflecting plates are such as to give best focus for any input selected. The third anode of the CRT is switched by S_{6D} to maintain focus.

A center-set 0- to 1-ma meter is used as a voltmeter to measure the difference in potential between either the horizontal or the vertical deflecting plates, its

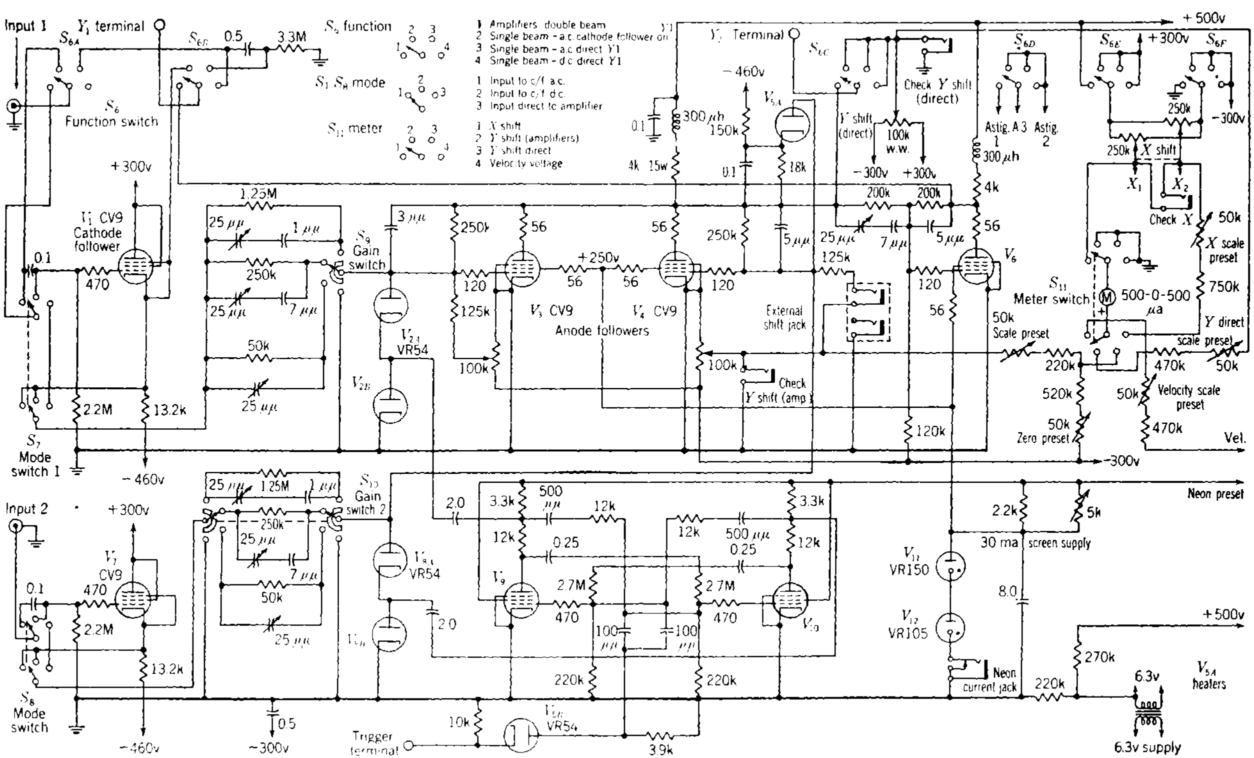


FIG. 18·50.—General-purpose monitor amplifier circuit.

function being selected by S_{11}. When the amplifiers are used, the meter is inserted in the grid-bias circuit of V_4 and measures input volts directly.

18·10. Other Radar Oscilloscopes.—Space limitations in this chapter do not permit the inclusion of all of the radar test oscilloscopes developed during the war period. Consequently, the instruments described in the preceding sections were selected in an attempt to present a representative group. There are other oscilloscopes that were produced in quantity and were rather widely used. A number of these will be mentioned briefly, and the references given may be consulted if additional information is desired.

The type Q oscilloscope (types TON-1GA and TON-1BL) is a small laboratory instrument that combines the functions of a synchroscope and a general-purpose oscilloscope. It uses a type 2AP1 cathode-ray tube with 25 to 5000 cps sawtooth sweeps and four triggered sweeps ranging from 4- to 1000-μsec duration. For external triggering, a positive or negative trigger of between 10 and 150 volts is required. The internal trigger generator covers the range of 185 to 4000 cps and puts out a 150-volt positive and 75-volt negative trigger. A 2Mc/sec oscillator is included for calibration of the 4-μsec triggered sweep. The single-stage signal amplifier has a bandwidth of 1.5 Mc/sec and a gain of 15 with a maximum input signal of 2.0 volts. This instrument was designed at the Radiation Laboratory and was manufactured by General Electronic Industries, Division of Auto Ordnance, Greenwich, Conn., and by the Browning Laboratories, Winchester, Mass. The operation and construction details are described in RL Report No. M-140 which contains the circuit diagrams A-6459-A and A-3975-A.

The TS-28/UPN is a test synchroscope, designed to check the range and coding circuits of beacon indicators, which has been supplied for use in servicing the YJ, YM, YK, and AN/CPN-3 radar beacons. A type 5CP1 cathode-ray tube is used with triggered sweep speeds ranging from 1 to 60 μsec/in. and free-running sawtooth sweep rates of 20 to 3000 cps. The triggered sweeps can be phased from -10 to $+100$ μsec with respect to the output trigger. Positive and negative output triggers having repetition rates of 330, 500, 1000, 2000, and 4000 cps are supplied, and markers having periods of 2, 10, and 25 μsec accurate to 1 per cent are generated for calibration of the sweeps. The two-stage video amplifier has a bandwidth of 5 Mc. This instrument was designed at the Radiation Laboratory and was manufactured by the Belmont Radio Corporation, Chicago, Ill. The preliminary instruction book is NAVSHIPS 900, 521-IB.

The TS239/UP is a general-purpose portable oscilloscope with type A presentation for use in the maintenance of radar systems. Its design specifications are the result of a fairly comprehensive survey of the requirements of a large number of radar systems, but volume production was not attained until a few months before the end of the war. A type 3BP1 cathode-ray tube is used with triggered sweeps having sweep rates in the range of 0.5 to 50,000 μsec per in. Any 10 per cent section of the sweep in the range of 10 to 50,000 μsec per in. can be expanded at least ten times. A variable delay controls the start of this expanded portion. The positive output trigger has an amplitude of 50 volts and a PRF of 400, 800, or 2000 cps. The video amplifier has a signal input range of 0.1 to 120 volts peak and is essentially flat from 20 cps to 5 Mc/sec. Calibration of this amplifier may be accomplished with an internally generated square wave signal (150 cps) which may be varied from 0.1 to 1 volt peak to peak. This unit was designed by the Navy Bureau of Ships and the Bell Telephone Laboratories, Inc.

The TS262/TPS-10 is a test oscilloscope designed for testing the AN/TPS-10 radar and AN/TPX-1 IFF equipment and is a part of the AN/MPM-13 test equipment. The indicator is a type 3PB1 cathode-ray tube with a variety of sawtooth and triggered sweeps. A delayed output trigger is provided so that the rise of all pulse waveforms in the IFF or radar modulator may be seen on the sweeps. The video amplifier has a gain of 100 and a bandwidth of 1.1 Mc/sec. An input to an r-f envelope detector is also provided for viewing magnetron pulses. A sawtooth voltage may be obtained from this instrument for sweeping an f-m test set. The construction of this oscilloscope is its most outstanding feature, as the entire unit is enclosed in a pressurized cylindrical container with operating controls brought out through panel seals. In addition, it is completely tropicalized and will operate at ambient temperatures of $-40°$ to $+120°$F so that satisfactory operation may be obtained under almost any climatic condition. This instrument was designed by the Radiation Laboratory and manufactured by the Harvey Radio Laboratories, Inc., Cambridge, Mass. Operational instructions are contained in the technical manuals for the AN/TPS-10: TM 11-1368, TM 11-1468, and TM 11-1568 published by the War Department.

18·11. Auxiliary Circuits.—A number of pieces of auxiliary equipment providing special functions have been designed for use with oscilloscopes not so equipped. Included in this category are synchronizers, marker generators, delay lines and circuits, and signal circuit devices such as probes, attenuators, and video amplifiers. Circuit designs for all of these can be found in the complete circuits of the oscilloscopes already described. The characteristics and references on a few more are included here for those wanting data on complete units designed for general purpose use.

The *Model G synchronizer* was designed to convert the standard commercial oscilloscopes into synchroscopes for laboratory use. Sweep writing speeds of 0.2 to 600 μsec are provided in nine steps when the synchronizer is used with an oscilloscope having a horizontal deflection factor of 40 volts per inch. These sweeps may be initiated by a positive or negative trigger having a minimum amplitude of 50 volts and a rise time of at least 0.5 μsec. Circuits providing a fixed trigger delay of 20 μsec and delay variable from 14 to 1200 μsec are included. The internally generated positive and negative triggers have an amplitude of about 200 volts, a rise time of 0.25 μsec, and a continuously variable PRF of 250 to 2500 cps. A pulsed oscillator having sinusoidal 100- and 500-kc/sec output voltages is used for sweep calibration. In addition, one tube is included for general-purpose use as an amplifier, inverter, or cathode follower. This unit is very flexible operationally, as the inputs and outputs of all circuits are brought out to panel jacks to be interconnected with jumper leads. The circuits diagram and operational details of this synchronizer are given in Vol. 22 of this series and are described in RL Report No. M-195A. It was produced by the Browning Laboratories, Inc., Winchester, Mass.

The *Model PH variable trigger-delay unit* was designed to provide a continuously variable delay between an input trigger and the output trigger supplied by the unit. A phantastron delay circuit using a type 6SA7 vacuum tube is used to cover two ranges of 10 to 600 μsec and 100 to 2400 μsec. It will accept a positive or negative input trigger of at least 30 volts amplitude and 0.3-μsec maximum rise time and generates delayed positive 75-volt triggers across 75 ohms impedance and 200-volt negative triggers across 4000 ohms. The rise time of the output triggers is 0.2 μsec. This unit was designed and a limited quantity produced at the Radiation Laboratory. It is described in RL Report No. 891 which includes the circuit diagram, RL drawing No. B-13967-A.

The *Model B*-8127 *sweep calibrator* is used to calibrate triggered sweeps in terms of time, to provide a basis for estimates of sweep linearity, and to generate timing markers for use in the analysis of other waveforms. The output of the calibrator consists of a series of narrow positive or negative pulses, having variable amplitudes of 0 to 40 volts. The first pulse generated immediately follows the synchronizing trigger. The time between successive pulses is 2.5, 10, 50, or 100 μsec which may be held to 0.2 per cent under normal conditions of operation. The number of markers produced is determined by an adjustable gating pulse which is continuously variable from 20- to 2500-μsec duration. Either the output of the internal trigger generator, having a PRF of 300 to 2000 cps, or an externally generated positive or negative trigger having a minimum amplitude of 50 volts, a rise time of at least 0.5 μsec, and a PRF of 200 to 4000 cps may be used to synchronize the marker generator.

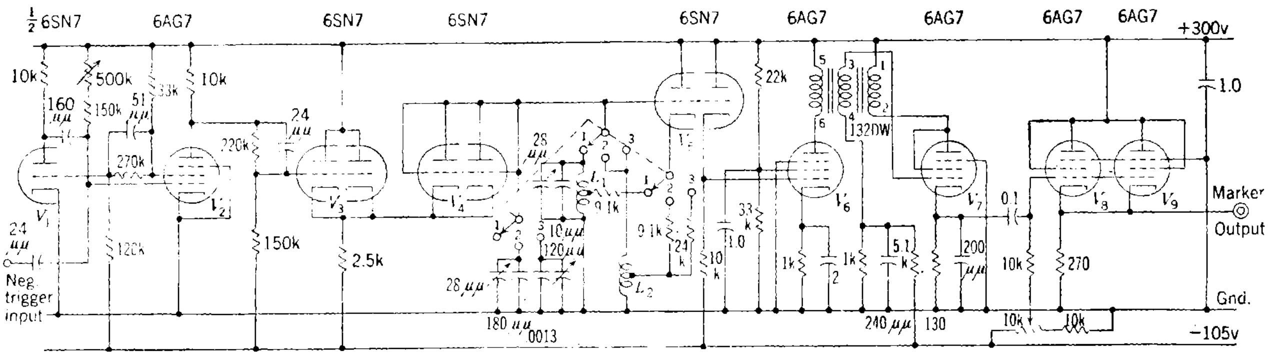


FIG. 18·51.—Triggered marker generator circuit.

Output triggers having amplitudes of 150 volts positive or 80 volts negative and rise times of 0.3 μsec are supplied for synchronizing external equipment. This unit was designed by the Radiation Laboratory and produced by the United Cinephone Corporation, Torrington, Conn. It is described in RL Report No. M-223.

The *Model C sweep calibrator* is somewhat simpler than the Model B-8127 just described but performs the same functions. It accepts an externally generated trigger having a minimum amplitude of 50 volts, a rise time of at least 0.3 μsec, and a PRF of 200 to 15,000 cps. The output is a series of positive calibration markers having time spacings of 0.3, 1, and 2.5 μsec for a maximum of 30 μsec following the input trigger. The marker amplitude is variable from 6 to 60 volts, and the output impedance is approximately 25 ohms. The circuit diagram of this unit is shown in Fig. 18·51. V_1 and V_2 form a triggered multivibrator whose output, a 30-μsec positive pulse, is coupled to the cathode follower V_3. The low-impedance pulse at the cathode of V_3 raises the cathode of the diode V_4 to a more positive potential than its plate, rapidly cutting off the current through this tube. This starts oscillation in the LC-circuit selected by the MARKER SPACING switch, and positive feedback from one section of V_5 maintains this oscillation at a constant amplitude. The sinusoidal output is peaked in V_6, firing the biased-off blocking oscillator V_7 on each cycle. Markers generated at the cathode of V_7 are coupled to the output jack through the parallel cathode followers V_8 and V_9. L_1 consists of 43 turns of No. 26 enameled wire on a 1-in. form with the tap 15 turns from the ground end, while L_2 has 82 turns of No. 28 enameled wire on a 1-in. form, tapped 26 turns from the ground end. This unit was designed and a limited quantity produced at the Radiation Laboratory.

A crystal-controlled range mark and trigger generator is shown in Fig. 18.52. This unit has a considerably higher frequency accuracy (about 0.01 per cent) than the triggered LC-controlled oscillators but must supply the synchronizing trigger to all equipment used with it. Crystals having frequencies in the range of 75 to 110 kc/sec may be used without component changes. Assuming a 100 kc/sec crystal, mixed 10- and 100-μsec markers are obtained at amplitudes of 20 and 40 volts respectively. Positive and negative output triggers have a maximum amplitude of 100 volts, a duration of 1 μsec, and a PRF of 200 to 2000 cps. Pips generated across the transformer winding in the cathode circuit of V_{1A} are used to synchronize the blocking oscillator V_{1B} at the crystal frequency. A current pulse from this stage is used to synchronize the 10/1 blocking oscillator V_{2B}, and the outputs of both stages are mixed and amplified in V_{2A} to obtain the marker output. A second blocking oscillator divider V_{3A} is synchronized with V_{2B} to obtain the trigger PRF. The positive output pulse from the tertiary winding of the transformer associated with V_{3A} is paraphased in V_{3B} to obtain the output triggers. This circuit was designed at the Radiation Laboratory, and a few of the units described were built there. Almost identical circuits were used in the AN/APS-15 radar range unit and are described in the instruction manual for that system.

The TS102/AP is a crystal-controlled range mark generator which was produced for field use. It is used with airborne radars; AN/APG-1, 2, and 13; AN/APQ-5, 7, and 10; and some ship radar systems. The output of a 327.8-kc/sec crystal oscillator is used to generate positive and negative markers having an amplitude of 0 to 30 volts, a duration

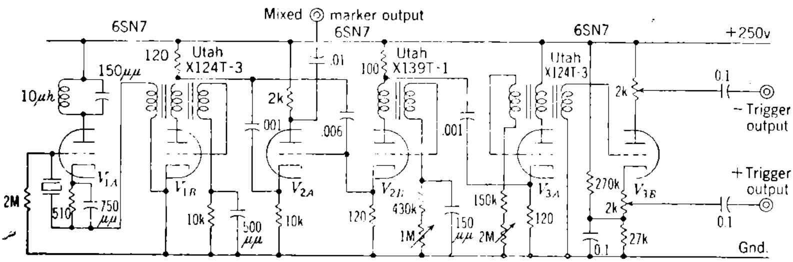


FIG. 18·52.—Crystal-controlled marker generator circuit.

of 0.4 μsec, and a spacing equivalent to 500 radar yards. These markers can be phased from 0° to 360° with respect to the output trigger. Both positive and negative output triggers having amplitudes of 50 volts, durations of 0.8 μsec, and a PRF of 400, 800, 1600, or 2000 cps are supplied from low-impedance outputs. A $\frac{1}{10}$-sec least-count manually operated stop watch for testing range rate calibration is also included. This instrument was designed by the Bell Telephone Laboratories and produced by the Western Electric Company, Inc. Western Electric Manual CO—AN08-35TS102-2 describes its operation and maintenance.

Several types of *electronic probes* have been developed to raise the input impedance of video amplifiers as discussed in Sec. 17·8. Two practical circuits are shown in Fig. 18·53. The first uses a type 6AK5 pentode as an inverter for negative pulses and a cathode follower for positive pulses to drive a 75-ohm line to the main video amplifier. The stage gain is approximately 0.1, and the bandwidth has been made as high as 70 Mc/sec. A resistance-capacitance divider reduces the input impedance to 1 megohm paralleled by 8 μμf.

The second probe is a cascade cathode follower with "bootstrapping" of both the grid and plate of the SD834 subminiature input triode. The effective input impedance is increased to about 20 megohms paralleled by 1 μμf. A larger tube than the 6Y6G should be employed if distortion of very fast negative pulses is to be avoided. Those components enclosed in the dotted lines are mounted in the head of the probe with the connections to the rest of the circuit made through a specially constructed 15-in. cable. The grid return and cathode leads are No. 30 Formex insulated wires running through two concentric Vinylite tubes. Over this is a $\frac{1}{4}$-in. braid which makes the connection to the plate of the SD834. The cable is completed by a final Vinylite outer tube for insulation.

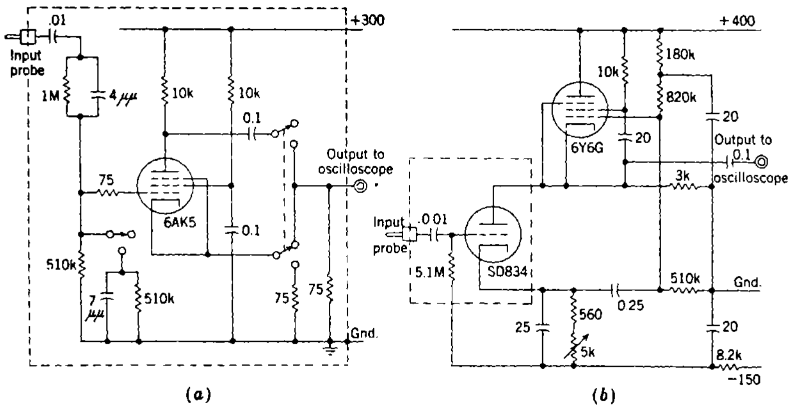


FIG. 18·53.—Probe circuits. (*a*) Single stage; (*b*) cascade cathode follower.

18·12. References.—A number of the manuals and reports listed in the present chapter have become available through the U.S. Department of Commerce. A selected bibliography of these and related reports follows.

1. PB-14842 *Handbook of Maintenance Instructions for TS51/APG-4 Test Set for AN/APG-4 and AN/APG-4X Aircraft Radar Equipments,* U.S. War and Navy Depts, Air Council United Kingdom, 1944.
2. PB-53894 *Handbook of Maintenance Instructions for Oscilloscope TS262A/TPS-10,* U.S. War Dept, 1946.
3. PB-15021 *Tentative Operating Instructions MIT RL Modified Type 102-A Test Set (Type 102A-1),* MIT RL Report M-123, 1943.
4. PB-16203 *Manual of Operation, Model #2 Synchroscope,* F. J. Gaffney, MIT RL Report M-112, 1942.
5. PB-32748 *Model P4-E Synchroscope and R-f Envelope Indicator,* MIT RL Report M-124; NDRC Report 14–440, 1943.
6. PB-2728 *Operating Instructions, RL Model 5 Synchroscope,* R. P. Abbenhouse, MIT RL Report M-212, 1945.
7. PB-2730 *Operating Instructions for Sweep Calibrator, Model B,* R. P. Abbenhouse, MIT RL Report M-212, 1944.
8. PB-15018 *Instructions for Operation of High Gain Video Amplifier for P4-E Synchroscope,* J. W. Severinghaus, MIT RL Report M-166, 1944.
9. PB-6839 *Instruction Manual: Revised Model P4 Synchroscope,* MIT RL Report M-126, 1944.

10. PB-8063 *Instruction Manual: Model P4 Synchroscope*, MIT RL Report M-118, 1943.
11. PB-16203 *Manual of Operation, Model #2 Synchroscope*, F. J. Gaffney, MIT RL Report M-112, 1942.
12. PB-2733 *Operating Instructions for Model G Synchronizer*, R. P. Abbenhouse and F. N. Gillette, MIT RL Report M-195A, 1945.
13. PB-3952 *Model II Calibrator*, Britton Chance, MIT RL Report 63–16, 1943.
14. PB-4013 *Instructions for Type "E" Self-synchronous Oscilloscope*, F. J. Gaffney, MIT RL Report M-109, 1942.
15. PB-14847 *Handbook of Maintenance Instructions for Oscilloscope TS34/AP*, U.S. War and Navy Depts, Air Council United Kingdom.
16 PB-16211 *Types TON-1GA (Type Q) and TON-1BL Oscilloscopes*, B. F. Wood, ed., MIT RL Report M-140, 1942.
17. PB-37898 *Instruction Book for Voltage Divider Probe, Type CAOR-62142*, U.S. Bureau of Ships, NAVSHIPS 900,299-LB, 1944.

PART V

THE DESIGN AND CONSTRUCTION OF ELECTRONIC APPARATUS

CHAPTER 19

THE DESIGN AND CONSTRUCTION OF ELECTRONIC APPARATUS

BY W. G. PROCTOR,[1] J. V. HOLDAM, JR., AND A. C. HUGHES, JR.

The present chapter contains some general considerations regarding the design and construction of electronic apparatus. Whereas the preceding parts of this volume have dealt with the functional design of various circuits, this part deals with practical design problems and construction methods.

The reader should appreciate that this chapter deals primarily with the design and construction of experimental and preproduction models of electronic apparatus. The experience on which these comments are based[2] was obtained at the Radiation Laboratory during World War II.

The design of equipment for military application has many problems not found in the design of equipment for commercial or laboratory use. Perhaps the most important of these is the extent to which components are standardized.

Another unique facet of designing equipment for military application is the existence of detailed performance specifications covering the operational performance of the equipment. From the military point of view the desirability of such specifications is obvious. The effect of such specifications on the designer, however, is quite drastic. The equipment designed for commercial use is often designed and then "sold" to customers. In the case of military specifications, however, the customer, i.e., the Armed Forces, dictates the performance. The difference rests, of course, on the fact that costs have little influence on equipment designed for military application whereas they are major factors in equipment designed for commercial application.

This chapter is divided into two topics: predesign, and design and construction. The former is a discussion of some of the more fundamental concepts, whereas the latter gives detailed recommendations of a

[1] Sections 19·1 to 19·8, inclusive, are by W. G. Proctor and J. V. Holdam, Jr.

[2] EDITORS' NOTE: To a greater extent than is true in other chapters of this volume the present chapter reflects the personal viewpoints of its authors on its somewhat controversial subject. It is presented as a background for further thought and discussion, rather than as a definitive treatment of the subject, and it is in this context that it should be approached by the reader.

modus operandi and contains some examples of the type of work with which the authors are most familiar. Considerable emphasis is given the special problems of lightweight design. For military applications, this has proved to be a very important part of the design problems.

19·1. Design Specifications.—In order to design a piece of electronic equipment properly one must have the following kinds of information: What is the equipment supposed to do; how much weight will be tolerated; what is the maximum power consumption; to what treatment will the equipment be subjected; and for such treatment, how do the available components and construction techniques respond? During the war this information was incorporated into specifications written by various branches of the Armed Service whose function it was to procure the electronic military equipment.

There are three classes of military specifications: general specifications which concern classes of equipment; equipment specifications which concern a specific equipment; and component specifications which concern the components used in building the equipment. The first two describe, among other things, conditions that equipments are likely to encounter and how tests are to be made to simulate those conditions. General specifications and equipment specifications are used by the design engineer as a goal toward which the design is pointed. The component specifications are used by the design engineer as a basis for the performance that he can expect from the individual components used in the equipment design.

Among other things the general specifications contain

1. Lists of applicable specifications and drawings covering components, spare parts, requirements for handbooks, etc.
2. The scope of the general specifications and equipment nomenclature.
3. Requirements as to materials and workmanship.
4. General requirements concerning interchangeability, safety devices, life, operating conditions, etc.
5. Detailed requirements applicable to the specific equipment.
6. Methods of test and inspection.
7. Packing instructions.
8. Notes that include the addresses of organizations from which the applicable or component specifications may be obtained.

One section of the general specifications, indicated by 5 above, is devoted to detailed requirements of the particular equipment. This section, the equipment specification, specifies performance, physical characteristics, and miscellaneous characteristics. The performance characteristics in the case of radar and associated equipment usually include such factors as detection range, coverage, scan rates, noise figure of the receiver, range accuracy, angular accuracy, etc. The specifications as to the physical characteristics include the allowable size of the components, how the complete system will be divided into components, the allowable weight of each component, the total weight of the whole system, the size of the antenna assembly, wave length, etc. The other specifications include such things as the permissible power requirements, required allowances for interconnection to other equipment, the accepted

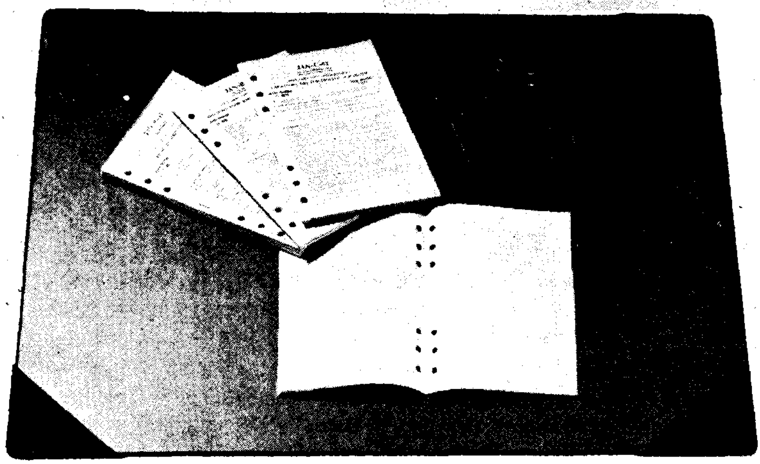


FIG. 19·1.—Sample specifications.

test equipment that must be used in testing, the minimum allowable power output, and requirements as to the ease of maintenance. The equipment specifications usually include waivers that nullify certain parts of the general specifications. For instance, requirements covering maximum altitude of operation for an airborne radar set, normally set at 50,000 ft, are waived and a new maximum altitude established if the radar set is not intended for high-altitude operation. The equipment specifications are almost always written in conjunction with the manufacturing company that plans to manufacture the equipment and generally reflect the manufacturer's feelings about the possibility of making equipment to meet the military requirements. During the war practically all of the equipment specifications were tentative; this allowed the manufacturers and designers considerable leeway in requesting modifications or waivers if the design turned out to be impossible or impractical.

The JAN[1] specifications probably comprise the most important list of component specifications. They cover practically all the elements of interest to the circuit designer; carbon, power, wire-wound, and precision wire-wound resistors; paper, electrolytic, mica, and mica substitute capacitors; transformers; inductors; impregnating materials; variable resistors; vacuum tubes; etc. The JAN specifications provide a very useful handbook of components, the prewar counterpart being the catalogue information available from the manufacturers. The advantage of having standardized specifications is that the same characteristics are specified for each component and more if not all of the characteristics are given for each component. A group of component specifications are shown in Fig. 19·1.

TABLE 19·1.—RANGE OF OPERATING CONDITIONS ENUMERATED IN GENERAL SPECIFICATIONS

	Airborne equipment	Shipborne equipment	Ground equipment
Temperature	−55° to +71.1°C operative; −55°C to +85°C nonoperative	0 to +50°C interior; −50° to +50°C exposed	−40° to +70°C
Relative humidity, %	95	95	95
Pressure	To 3–5 in. Hg (50,000-ft altitude)	. . . .	. . . .
Voltage fluctuation	Aircraft d-c supply. 22 to 30 volts	Primary voltage ±10%	Primary voltage ±10%
Vibration	10 to 500 cps Max. amp 0.03 in.	Ship: 5 to 30 cps Max. amp $\frac{1}{16}$ in. Speed craft: 10 to 300 cps.	5–250 cps
Shock	10 g	100–200 g	25–50 g

19·2. Use of Specifications.—The general specification has a very profound influence on design. Perhaps the greatest influence is exerted by that part of the general specification which sets forth the extremes in operating variables such as temperature, humidity, and pressure. Table 19·1 shows the different expected extremes in the more important variables for the three major classifications of radar equipment, i.e., airborne,

[1] Joint Army-Navy (JAN) specifications are the result of coordinated effort toward standardization on the part of representatives from the Armed Forces, Government procurement and production agencies, and interested industries.

shipborne, and ground. Information in the general specification, as exemplified in Table 19·1, has the effect on the design of immediately limiting the possible circuit designs, construction techniques, and components.

The design restrictions imposed by the general specification rarely come as a surprise to the design engineer, since considerable thought has usually been placed on the design problems before specifications are written.

The equipment specifications determine the exact circuit design. It is this part of the specifications which is subject to the greatest amount of discussion. In general, the using service tries to write the specifications so as to squeeze the maximum possible performance out of the equipment being built, whereas the design engineer attempts to have the equipment specifications written to give him the greatest leeway in circuit design. As has been pointed out in the preceding section, the equipment specifications rarely become final until the equipment is in production. The design stage is usually based on tentative equipment specifications which are constantly being changed as a result of interchange of ideas and opinions between the design engineers and the procuring Services. As far as radar equipment is concerned, equipment specifications usually contain considerable leeway as regards performance. This arises from an inherent characteristic of all radar sets, that performance depends not only upon the quality of circuit design but also upon the quality of maintenance and skill of the operator.

Component specifications are the designer's building blocks in design just as the components themselves are the building blocks in construction. Usually, each of the variable conditions, specified in the general specifications under which the over-all equipment must operate, will cause some variation in the characteristics of each component. The designer uses the component specifications to select the component that has variations under those conditions which do not limit the operation of the equipment. This is the meat of design once the general circuit has been evolved. It may well turn out that the general specifications influence the design to a much greater extent than simply determining component selection. For instance, for certain types of mechanical construction some of the possible operating conditions may be mutually exclusive to satisfactory operation for *any* available component. If such a situation exists, it is then necessary to modify the condition that the component has to meet by special mechanical or electrical arrangement of the circuit. An excellent example of this situation is the design of modulators used in radar systems. The practice of pressurizing such units has become almost universal. This has been necessary in order to meet the altitude specification and the weight specification in airborne radar systems. By making the modulator pressurized, the components are not subjected to the low barometric pressure even though the airplane is flying at very high altitude; furthermore, the components are not subjected to high humidity conditions.

Since there is only a small number of such tricks that can be employed by the designer, it is generally necessary to select components that will meet the general specifications. The following is an example of the type of information included in the component specification in the JAN handbook. Since large numbers of resistors are used in electronic equipment, a 1-megohm carbon resistor is taken as an example.

1. Temperature coefficient. A 1.0-megohm carbon resistor (at 25°C) may have any value between 1.0 and 1.52 megohms at −55°C; it may have any value between 1.36 and 0.74 megohm at 105°C.
2. Derating. A carbon resistor is expected to meet its load-life requirements when dissipating its full rated wattage at an ambient of 40°C, 50 per cent of its rated wattage at an ambient of 70°C, or 10 per cent of its rated wattage at an ambient of 94°C.
3. Load life. Resistors are required to operate at their rated wattages at an ambient of 40°C without suffering permanent change in resistance of over 10 per cent. At an ambient of 85°C, this requirement is 200 hr.
4. Overload. Resistors are required to stand voltages of 2.5 times rated continuous working voltages for 5 sec without suffering a permanent change in resistance over 5 per cent. This is, of course 6.25 times the rated wattage.
5. Voltage coefficient. A $\frac{1}{2}$-watt resistor may measure 3.5 per cent greater at 100 volts than at 10 volts. (Hence distortion may be introduced with resistor dividers used with large-voltage waveforms.)
6. Security of terminals. Resistor terminals are expected to withstand a 5-lb pull without sustaining mechanical injury.
7. Exposure to humidity. After an exposure to 95 per cent humidity at 40°C for 250 hr, the change in resistance of a resistor should not exceed 10 per cent. (Hence, a few per cent change is possible as a consequence of a few weeks' shelf life in humid regions.)

Some components, principally tubes, are not available with different characteristics. For these components, the designer simply has to face the manufacturing tolerances and design his circuit so that it will operate properly with either low-limit or high-limit components. An example of this type of component is given in Table 19·2, which lists some of the characteristics of several common vacuum tubes. A perusal of this table will make it immediately obvious that one of the principal tasks facing the designer is the reconciliation of the wide tolerances that he must expect from the vacuum tubes with the performance requirements set forth in the general and equipment specifications.

It sometimes happens that some important characteristic of a component is not mentioned or specified in the available component specifications. In such a situation, it is necessary for the design engineer to establish for himself the probable limits of this characteristic for available components and select a component that best meets his need. For instance, the JAN specifications of carbon resistors do not cover the variation of resistance with frequency or the amount of inductance in an average resistor.

TABLE 19·2.—TYPICAL JAN SPECIFICATIONS FOR VACUUM TUBES

Tube	Plate current at specified fixed bias, ma		% deviation from mean	Transconductance at fixed voltages, μ mhos		% deviation from mean
	Min.	Max.		Min.	Max.	
6C4	6.5	14.5	±38	2500	4000	±23
6SJ7	2.0	4.0	±33	1325	1975	±20
6SK7	6.5	12.0	±30	1600	2400	±20
6SL7	1.4	3.2	±39	1200	2000	±25
6SN7	5.5	12.5	±39	2075	3125	±20
6V6	3.3	5.7	±27	3000	5200	±22

19.3. Fundamentals of Temperature-rise Analysis.—One of the most serious problems that the design engineer must face is proper mechanical design to facilitate cooling of components. The data given in Table 19·1 indicates that equipment built to military specifications must be designed to operate over very wide variations in ambient temperature. On the other hand, the data given in Sec. 19·2 on the performance of a standard resistor component as a function of variations in ambient temperature point up the difficulties in designing equipment with components whose characteristics vary so widely with changes in ambient temperature. Certainly for airborne equipment, the single most difficult problem arises from the wide variation in pressure and temperature. The difficulty usually arises from the fact that the equipment is adjusted under high-temperature conditions (on the ground) whereas it is expected to operate, without further adjustment, under low-temperature conditions (in the air). Although the equipment has its principal operating use under medium- or low-temperature conditions, it must be designed so that it will not "burn up" when operated under high-temperature conditions.

Since large temperature variations do so much to restrict the design, considerable effort has been directed to developing efficient means of cooling equipment. This section treats in an elementary fashion the information that the designer needs in order to plan the cooling in the system; basic heat transfer formulas are given, and the cooling of closed and open units is discussed.[1]

Correlation between calculations based on the formulas given here and the physical tests is good only if the physical configuration of the heat-dissipating object is simple. Consequently, the formulas are of principal advantage only in showing the basic limitations of heat transfer so that the physical configuration in the unit can be adjusted for maximum efficiency. It is almost always necessary to confirm any analysis by experiment.

Basic Heat-transfer Formulas.—Heat is transferred in three familiar ways: conduction, radiation, and convection. Expressions are given for these three heat-transfer methods, more because they form the basis for subsequent discussion than for any use that might be made of them directly.[2] Only the first two are based upon physical law; the expressions for convection were derived empirically. Unmodified air is assumed to be at 20°C, 50 per cent humidity, and standard atmospheric pressure.

1. *Conduction.*—The power dissipated by conduction is

$$P = AK\frac{\Delta T}{d} \qquad \text{watts,} \tag{1}$$

where A is the area of the heat conducting region in square inches normal to the direction of heat flow, ΔT is the temperature difference existing between two points d in. apart, and K is a constant of heat conductivity which differs with the material: copper 9.0, iron 2.0, aluminum 3.3, carbon steel 1.1, rubber 0.004, and cloth or felt 0.0016. Thus for steel about 1 watt is conducted through 1 sq. in. of contact surface for points 1 in. apart having a temperature difference of 1°C. Note that *components which are required to dissipate large amounts of heat should be mounted so that continuous metallic contact is made to a large metallic surface, such as the frame.*

2. *Radiation.*—The power dissipated by radiation is

$$P = 37e10^{-12}[(T_a + \Delta T)^4 - \mathrm{T}_3{}^4] \qquad \text{watts/in}^2 \tag{2}$$

where T_a is the absolute ambient temperature, ΔT is the temperature difference in degrees centigrade between the radiating body and the surrounding objects, and e is an emissivity constant that varies

[1] For analysis of complicated heat-transfer problems, volumes such as McAdams, *Heat Transmission*, McGraw-Hill, 1942, should be consulted.

[2] These basic formulas for heat transfer were taken from *Magnetic Circuits and Transformers* by the MIT Staff in Electrical Engineering.

for different materials, typical values being ideal black body 1.0, lamp black 0.98, rough insulating materials 0.90, oxidized iron 0.75, oxidized copper 0.65, polished aluminum 0.005, and polished silver (mirror), 0.0025.

An approximate form of Eq. (2), which gives a good approximation if ΔT is small, is

$$P = 37e10^{-12}4T_a^3\,\Delta T. \tag{3}$$

For an ideal black body with a temperature difference of 1° above its surroundings, at 50°C the power radiated is 0.005 watts per square inch of surface.

3. *Convection in Still Air.*—The power dissipated by convection in still air is

$$P = 0.0016(\Delta T)^{1.25} \qquad \text{watts/in}^2 \tag{4}$$

for the upper side of a horizontal surface,

$$P = 0.0012(\Delta T)^{1.25} \qquad \text{watts/in}^2 \tag{5}$$

for a vertical surface, and

$$P = 0.0008(\Delta T)^{1.25} \qquad \text{watts/in}^2 \tag{6}$$

for the bottom side of a horizontal surface, where ΔT is the temperature difference in degrees centigrade between the surface and the air about it. A 1-in. cube having a temperature 1°C higher than the surrounding air loses heat by convection at the rate of 0.0072 watt.

Cooling a Sealed Unit.—General empirical formulas relating the heat transfer from a heat generator inside a sealed unit to the outside atmosphere are not too satisfactory, because the efficiency of heat transfer is a sensitive function of the geometrical configuration of the heat-producing components and the sealed housing. The geometry to be considered is a heat source in air surrounded by an airtight metallic container also in air. This treatment does not include conduction of heat to the wall of the container by metallic contact; the effect of such conduction, however, can be estimated from Eq. (1). In any practical system, it is necessary to consider the flow of heat from several components through the housing to the surrounding atmosphere. This can be done approximately by summing the results of calculation of the heat flow from each separate component.

The basic formula for heat transfer under the conditions specified above is

$$P = UA\,\Delta T \qquad \text{watts,} \tag{7}$$

where U is a coefficient of heat flow in watts per square inch of exposed surface per degree centigrade, A is the exposed area in square inches, and ΔT is the temperature difference in degrees centigrade of the air inside and outside the container. The problem, of course, is evaluating U for the components under investigation.

In a practical design U is usually the only variable; P is simply the total amount of power in watts generated inside the unit. The temperature difference ΔT, in degrees centigrade, is the difference between the maximum operating ambient specified in the general specifications and the maximum operating ambient for the most sensitive component specified in the component specifications. For instance, most components are not guaranteed by the manufacturer for operation at temperatures above 100°C, but most systems must operate at an ambient temperature of at least 70°C, so ΔT is usually of the order of 30°C.

For these conditions, the heat emission by radiation can be neglected; and since the walls of the containers are usually thin metal, the resistance to the conduction of heat through the metal is negligible. The principal resistance to heat flow is caused by the air films on the surface of the component and on the surface of the housing. The value of the over-all coefficient of heat transmission is the sum of the resistances to its flow and is expressed as a conductivity by the formula

$$U = \frac{1}{\dfrac{R}{h_i} + \dfrac{X}{K} + \dfrac{1}{zh_0}}, \tag{8}$$

where U is the over-all coefficient of heat transfer in watts per square inch per degree centigrade, the surface is the external surface, and the temperature is the mean temperature difference between the air inside the cabinet and outside the cabinet; h_i is the internal film coefficient of heat transfer in watts per square inch per degree centigrade, the surface is the internal surface, and the temperature is the mean temperature difference between the inside air and the inside surface; h_0 is the external film coefficient of heat transfer in watts per square inch per degree centigrade, the surface is the external surface, and the temperature is the mean temperature difference between the outside air and the external surface; R is the ratio of the external surface area to the internal surface area and multiplies the value of the internal film coefficient so that the outside surface may be used in calculations; z is the efficiency of the external surface and differs from one only when fins or other extended surfaces are used. The thickness of the wall of the cabinet in inches is represented by X, and K is the coefficient of heat conductivity of the material of the cabinet.

With most metal cabinets, the term X/K is negligible, and R and z are unity. When this is true, Eq. (8) becomes

$$U = \frac{1}{\dfrac{1}{h_i} + \dfrac{1}{h_0}}. \tag{9}$$

Considerable experimental data at the Radiation Laboratory have established certain values of h for conditions approximating those found in the metal housings used in practice. For the following configurations and air conditions, h has the values indicated.

1. The value of h for any surface on the inside or outside of a closed unit may be taken to be 0.007 watt/in^2/°C when there is no forced circulation.
2. Using a No. 2, 15-watt blower-motor combination, supplying 15 cu ft of air per minute at a nozzle velocity of 3500 ft/min on the inside of a cylindrical container 9 in. in diameter and 15 in. long, the value found for h_i was 0.025 watt/sq in./°C.

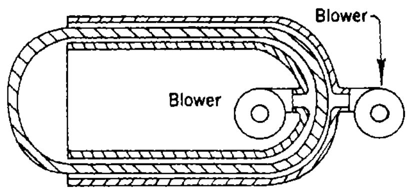


FIG. 19·2.—Triple-walled container.

3. The same blower, directing air at one end of the same unit on the outside, gave a value for h_0 of 0.015 watt/sq in./°C.
4. With small amounts of forced convection parallel to the surface, h approaches 0.02 watt/sq in./°C. From this point on, the value increases about linearly to 0.055 watt/sq in./°C with an air velocity of 42 ft/sec, measured where the presence of the surface does not affect the velocity. A way of achieving such a flow over the surfaces of a unit is illustrated in Fig. 19·2. Only one of the walls completely encloses the unit; the others are only partial walls used to restrict the flow of air to a broad, flat stream.

The figures given are approximations only; they can be reproduced reasonably well provided the physical arrangements have the proper configuration. Also, the coefficients may be deceptive, since, as the capacity of the inside blower is increased, the additional wattage required for the blower motor must be added to the internal heat load. In most equipment, there is an optimum rate of heat transfer because increasing the average heat transfer may mean a greater variation of the temperatures in various parts of the container. The total heat transfer is based upon the mean temperature difference between the air in contact with the inside and outside surfaces.

In Fig. 19·2 it is apparent that different film coefficients apply to different regions of the unit. The area of the surface may be divided up into different regions where judgment or simple experiment (as studying air flow by the drift of small bits of tissue paper) indicates different values of h. The heat carried away from each of these regions is then summed.

Cooling an Open Cabinet with Changing Air.—When it is not necessary to have a sealed cabinet, heat may be carried away by drawing air from the outside, circulating it through the cabinet, and discharging it again on the outside. The specific heat of air at 50 per cent relative humidity is 0.43 Btu/lb/°C. At 20°C and atmospheric pressure this is 0.57 watt-min/ft^3/°C. As the relative humidity increases to 100 per cent, the specific heat of air increases to a maximum value of 0.44; hence, humidity is neglected in the calculations.

The heat loss is

$$P = N\,\Delta TV \qquad \text{watts,} \tag{10}$$

where ΔT is the mean temperature difference between the air entering the unit and the air leaving the unit in degrees centigrade, N is a coefficient for the specific heat of air obtained by multiplying 0.57 watt-min/ft^3/°C by the ratio of 290°A to the absolute temperature of either the input or output air temperature (since the density changes with temperature), and V, in cubic feet per minute, is the rate at which air is being exchanged through the unit.

For positive circulation, fans or blowers must be used. The blower may be located at the intake, pushing the air through the cabinet; it may be located at the discharge; or two blowers may be used. Theoretically, it is more efficient to have the blower handle the coolest air. Owing to interferences that cause eddy currents, however, a better circulation may be obtained by exhausting the warmer air. As a general rule, the path of the air stream should be vertically upward to obtain the benefit of the gravity circulation; ducts or baffles are useful in obtaining adequate circulation. Blowers, turned by a motor of given rating, are rated for certain air volumes which remain independent of small back pressures that are obtained as air is blown through the unit.

The incorporation of filters into the intake port is common practice when the equipment is to be used in dusty atmospheres. A typical filter is about $\frac{3}{4}$ in. thick and made up of many layers of fine metal screening covered with oil. Such filters have a recommended capacity of about 2 cu ft of air per minute per square inch. The resistance of the filters to the flow of air constitutes practically all of the resistance to air flow in the unit. Consequently, it is necessary to know the filter-pressure air-velocity characteristics before the proper blower and motor can be determined.

Although the total amount of heat being removed from the unit can be determined by measuring the temperature difference at the input and exhaust ports, separate experiments are necessary to determine whether or not hot spots exist in the unit. It is often possible to equalize the temperature inside the unit by appropriate use of baffles; however, it is occasionally necessary to use separate small blowers directing air upon the hot components. The same principles are involved in cooling a separate component as in cooling a complete unit.

If the heat dissipated inside the unit is not too great, gravity cooling may be employed. This has the obvious advantage of requiring no fans which are, themselves, a source of maintenance trouble. The temperature rise can be calculated using Eqs. (7) and (8) and using the value for U given for natural convection. The use of louvers and slots reduces the temperature rise to about two-thirds the value calculated from Eqs. (7) and (8).

Testing Procedures.—As previously indicated, it is very difficult to predict accurately the temperature rise either in the unit as a whole or in any particular section. As a result, heat runs and tests are necessary to determine the exact performance. The most satisfactory method of testing is by the use of thermocouples. By using several thermocouples connected through a multiple selector switch, it is possible to measure the temperature variations within the unit during one heat run. Techniques that are less accurate but often useful employ the use of crayons, paints, and pills that change color as a function of temperature. These techniques are particularly valuable during the trial-and-error stage of adjusting mechanical baffles to eliminate hot spots on a chassis.

In setting up tests, several factors affect the relation between the test results and actual operation.

1. For a given power input and cooling arrangement, the temperature rise above ambient varies directly with the ambient. Over an ambient range of 10° to 50°C the rise above ambient increases approximately 10 per cent. (This is due to the change in the density of the air.)
2. Type tests of complete sets under specified low and high ambient conditions require large test chambers, in which a high rate of air circulation must be maintained in order to maintain the required temperature and humidity conditions. This means that the heat dissipation will be greater than in still air conditions.
3. When the cooling air is supplied to airborne equipment by air-scoops, the decrease in density due to altitude is offset by the decrease in ambient up to 40,000 ft. Such equipment must be supplied with auxiliary cooling, however, for ground operation.

19·4. Lightweight Apparatus.—During the war, considerable effort was placed on the development of lightweight apparatus. The development was carried out at considerable cost in engineering and design time both on the part of system design engineers and on the part of the components manufacturers. The emphasis for lightweight construction and design was originally placed on equipment to be used by the Air Forces. It later became apparent that all electronic equipment should be lightweight, since its utility is greatly enhanced if it is air-transportable.

The importance of saving a few pounds in the design of an equipment cannot be overlooked. For instance, it is estimated that a saving of 1 lb in the net weight of a commercial aircraft is worth approximately $2000 to the air lines during the life of the aircraft. Stated another way, if enough weight and space can be saved by proper design of equipment so that an additional seat can be added to a commercial aircraft, the saving is worth roughly $38,000 to the air line.

By directing so much attention to the problem, great strides have been made in reducing the weight of electronic equipment. This has been accomplished by coordinated effort on the part of the design engineers and on the part of the designers and manufacturers of components.

One of the most serious limitations to lightweight construction is the inherent ruggedness of the equipment as required by military specifications. Early experience showed, unfortunately, that breakage is one of the most serious problems connected with supplying combat units. As a result of this experience, military specifications now require that the equipment be very rugged. In the postwar development of commercial applications many of these specifications can be ignored because the equipment will be handled more carefully; nevertheless, war-learned lessons in building electronic equipment for rough treatment should not be quickly forgotten.

19·5. Minimum-weight Design.—The comments that follow are concerned with weight savings by proper circuit design; a later section contains information on weight-saving by construction techniques. The discussion presumes that the basic circuits have been satisfactorily worked out. The circuit components and considerations that hold the most promise for weight-saving are power consumption, power supply, tube selection, tube operating conditions, and component selection. Obviously these are related, but a detailed consideration of each points out the salient factors to be investigated.

Power Consumption.—It is often the case that component size is determined by the total amount of power consumed rather than the "compressibility" of the circuit. This is so because the power generated in the unit must be dissipated to prevent the components from operating at excessive temperatures. The four principal sources of heat in a vacuum-tube circuit are tube filament, plate dissipation, load dissipation, and voltage divider dissipation. These sources of heat can be either eliminated or cut to the minimum. It is usually desirable to determine the minimum plate voltage supply for each tube. It will often happen that enough tubes operate at the same voltage to warrant a special voltage supply. Cathode followers, input amplifiers, and other power-consuming circuits that do not contribute materially to the circuit operation are eliminated if at all possible. This can often be done by minor rearrangement in the circuit and by careful proportioning of the circuit into the various major units.

Wherever possible, tubes are eliminated, and those which are used are chosen for low heater power. For instance, diodes can sometimes be replaced by crystal or electrolytic detectors. The use of double tubes generally results in less power dissipation than the use of single tubes.

Regulated voltage is always costly in terms of transformer and filter weight and size; consequently, each circuit is analyzed to determine the exact amount of regulation required. Considerable power is saved if circuits that do not require regulation are not run on the regulated power supply.

Power Supplies.—It is often the case that the heaviest component in a unit is the power transformer. Maximum design efficiency is achieved only by careful analysis of the transformer requirements and by designing a special transformer to meet those requirements. As indicated above, taps on the transformer can be used to supply medium voltage buses, thus eliminating the necessity of dropping resistors from a high-voltage bus. This is particularly applicable to circuits that use many pentodes, since the screen potential is usually constant. Supplying the screen voltage by means of a separate voltage bus eliminates the power loss in a dropping resistor and, being low impedance, reduces the number of screen bypass condensers required. For many applications it is possible to use the primary voltage supplied without passing it through a transformer. For instance, for electronic equipment designed to operate in aircraft, considerable saving is effected by running the filaments in series parallel from the 28-volt d-c bus. If filament stabilization is required, nonlinear series elements can be incorporated into the circuit without requiring the transformers to supply the wasted power. For some applications, the use of a dynamotor to supply the high voltage is more efficient than converting direct to alternating current in a rotary converter and stepping up the voltage in a power transformer and rectifier circuit.

Component Selection.—Volume 17 contains descriptions and illustrations of most of the important components that have been developed during the last three or four years. In almost every case, the component has been reduced in size and weight, and the reliability has been improved.

Component size affects the final weight of a unit in ways that are not immediately obvious.

In most cases, the reduced size and weight have been effected by the development of new insulating materials and sealing compounds (for transformers and condensers) and by ignoring preconceived ideas of size (miniature and subminiature tubes). As indicated above, the development of these smaller and lighter components has been very costly, but some of the cost has been amortized so that the components are now available for commercial use at more reasonable prices.

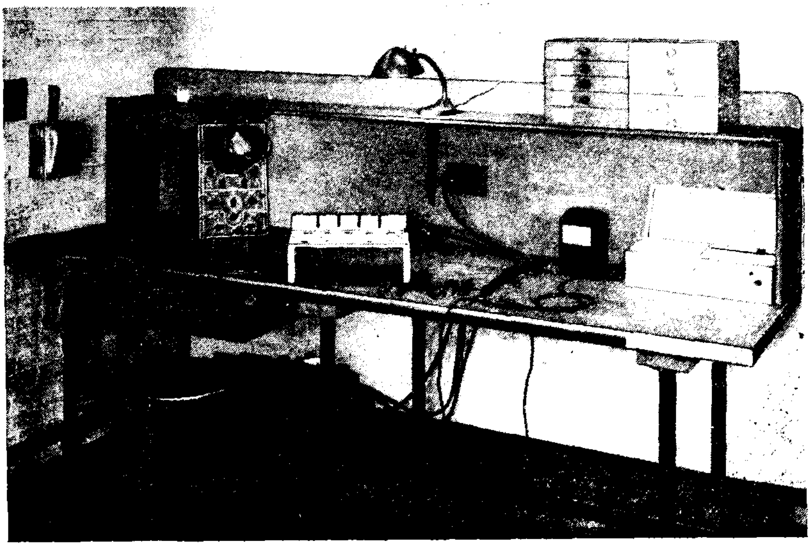


FIG. 19·3.—Typical work space.

19·6. The Design Process.—There is no guaranteed formula for generating new ideas, and this chapter does not purport to give one. It is assumed that the designer of new equipment has the necessary creative ability. This section gives some of the design techniques that the authors have found to be useful.

Figure 19·3 shows a typical work space assigned to a circuit designer, and Figs. 19·4 and 19·5 show typical examples of electrical mockup chassis (ignominiously called "breadboards"). Results are produced only when such physical arrangements are augmented by skilled specialists. Such personnel are not included in the list of illustrations, but they rank far above physical arrangements in importance.

The design process outlined in this section is idealized for the sake

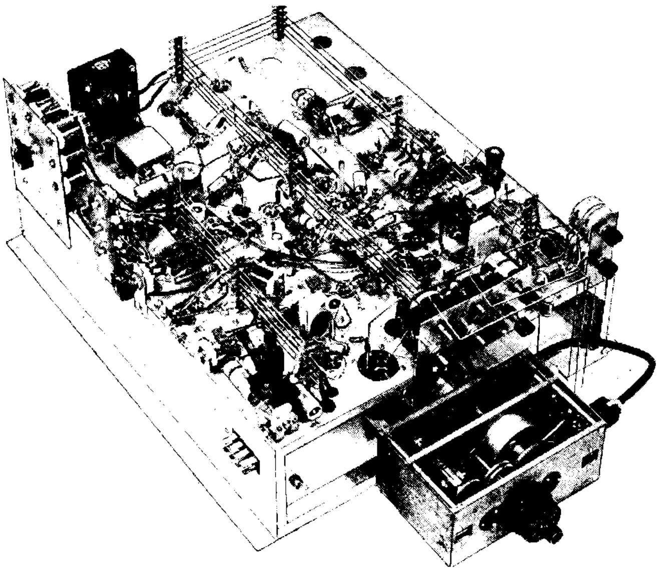


FIG. 19·4.—Metal electrical mockup chassis.

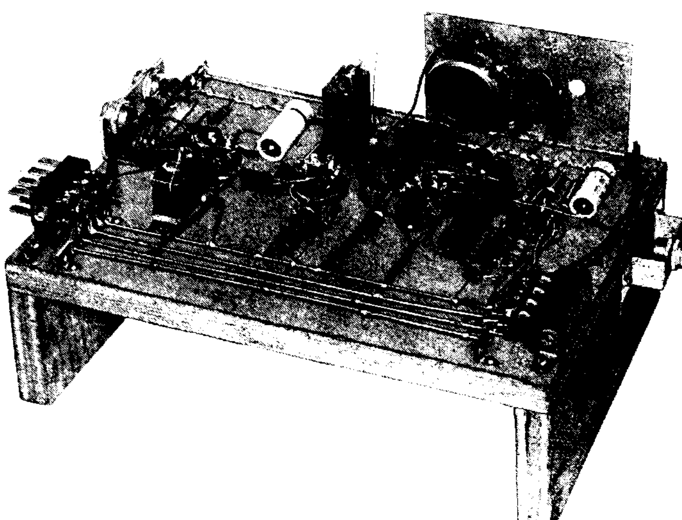


FIG. 19·5.—Wooden electrical mockup chassis.

of generality. It assumes that the design engineer has at hand all of the information discussed in the preceding sections and has the physical facilities and supporting personnel necessary to carry out the separate steps. Each step in the design process is indicated by a numbered paragraph. It is probably fair to say that no system or part of a system has ever been designed by the process outlined below. The urgency of war often established a time schedule that was incompatible with carrying out the complete process. Experience has shown, however, that while some of the steps may be combined, none of them can be omitted without risk of unfortunate results.

The problem that is considered is the design of a functional part of a complete system, such as a modulator, synchronizer, or range unit of a complete radar system. For very complex systems, the responsibility of design may be subdivided even further; this requires greater coordination among the designers working on each separate part.

1. The first step is to make a block diagram of the complete system to be designed. Such diagrams are very important so that each designer working on the individual units knows exactly how his unit fits into the whole system. Usually, the system is broken down into functional units; but as the units are designed, minor changes may be made. One of the most important benefits of a system block diagram is to specify the various inputs and outputs of each block to enable the designer of each block to know what he has in the way of inputs and what he has to supply in the way of outputs. The process of writing down the inputs and outputs of each block usually points up places where over-all efficiency can be improved by not following strictly functional subdivisions. For example, a functional division of a radar system that includes a receiver and a range unit would put the receiver in one unit and the range unit in another. It is generally more efficient, however, to put the final video stages of the receiver in the range unit, thus allowing interconnection between the receiver and the range unit at low level. This results in considerable saving of power.

2. After the complete system is broken down into blocks, the design engineer in charge of developing a particular block makes a block diagram of his unit. In this breakdown, the complex unit is divided into convenient components each of the same order of complexity which, in turn, may be broken down into single-function circuits that initiate or operate upon waveforms. The purpose of the block diagram is to organize the job and to clarify the relationship that exists among different parts of the circuit. Waveforms, voltage levels, impedance levels, and timing diagrams clarify the unit whenever they are known.

3. Having specified the complete functions and known operating conditions of each block, the designer fills in the blocks with circuit dia-

grams. This is usually done by one of two methods or a combination of two methods. In breaking down the unit into functional circuits, the designer makes use of his previous experience and knowledge of the field of electronic design; consequently, many of the blocks can be filled in by circuits with which the designer has had previous experience. The circuits will rarely be in a finished form, since they have not been previously applied to the particular application that the designer faces. Previously developed circuits, however, are an excellent start toward filling out the complete unit. The second method is based upon the designer's fundamental knowledge of the operation of vacuum tubes and electronic components. Knowing how these components behave and knowing what the circuit is supposed to do enable the designer to lay out a rough circuit diagram. At this point in the design, the unit consists of circuits whose general performance is known but that have never been tried in the particular application and circuits that have been invented on the basis of knowledge of the desired functional performance. The former, having been previously tested, have their principal constants already fixed, whereas the latter must have the constants determined.

4. A theoretical analysis of the circuit is useful at this point. This consists of reviewing the development of the proved circuit to establish a firm basis for the minor adjustments called for. An independent analysis is made of the new circuits to establish the probable range of circuit constants and the theoretical performance. Such analysis usually points to alternative methods of performing some of the functions and is used as a basis for experimental tests which follow.

5. Since the usual assumptions made in theoretical circuit analysis are quite broad, experimental checks are necessary. Often the designer will interrupt analysis to build a circuit and test it. On the other hand, when an experimental circuit exhibits unexpected characteristics, the designer usually reverts to analysis for explanation of the phenomena. In any case, experimental checks of the various circuits are necessary. This phase of design is very important, as it is a basis for the establishment of the variable parameters in the circuit. As a rule, the circuits are tested in as small units as possible to reduce the number of parameters and to facilitate a detailed examination of each parameter.

6. After the parameters and operating conditions of the new circuits have been satisfactorily established and the old circuits have been reworked to fit into the rest of the design, the complete unit is wired up on an electrical mockup chassis. Unforeseen difficulties usually arise when the whole unit (less power supply) is tried out for the first time. It is generally necessary to change some of the constants and perhaps even add new elements. Since the unit in this stage of development has little physical resemblance to the final unit, many important char-

acteristics cannot be tested. If the initial checks indicate that general operation is satisfactory, the remaining steps in the design process can be undertaken. If the unit as a whole does not check satisfactorily, then it is necessary to repeat some of the steps already carried out.

Developing a circuit up to this point is often considered the principal problem; certainly it is the most interesting to the designer. Satisfactory operation on an electrical mockup chassis, however, does not preclude unsatisfactory operation of the finished product. It is very important that the remaining steps in the design process be faithfully followed.

7. The next step in the design process is to test the whole unit to establish the power requirements, the required voltage stability, and the heat dissipation, and to find circuit elements critical to physical configurations. In general, the initial determination of these characteristics will be very disappointing, and it then becomes the job of the designer to work over the circuit to establish efficient conditions for each of these variables. As was pointed out in Secs. 19·4 and 19·5, the weight of the unit is closely allied to the amount of power that it consumes. The power requirements must therefore be cut to the minimum.

Manufacturing difficulty is closely allied to the tolerances required in the components. Operating stability is closely allied to component tolerances and optimum subassembly design; stability can usually be achieved by using precision components or by matching components. In many instances it is better, from the manufacturing point of view, to match components in a compact subassembly design than it is to use precision components. Such techniques also increase the serviceability of the equipment by making possible easy and rapid changing of subassembly units.

As pointed out in Secs. 19·4 and 19·5, considerable saving in weight is accomplished by reducing the requirements on voltage stability. Regulated power supplies are necessarily less efficient and heavier than unregulated power supplies. A careful test of the required voltage stability permits the power supply to be regulated no better than is necessary, to avoid unnecessary weight.

8. The information gained in these tests is coalesced with a general circuit review to check for wasted power, unnecessary tubes and components, proper interconnection to other units; to design the power supply and subassemblies; to decide on a satisfactory method for testing, maintenance, and calibration; and to make up a tentative parts list. Best results are usually obtained by working on these problems concurrently. Some of these considerations receive more emphasis than others, depending upon the use to which the equipment is to be put. For airborne equipment, particular emphasis is put on reducing weight and

volume; for other applications, these characteristics assume less importance. As work progresses on this step, the various subassembly designs are fabricated so that the effect on the other characteristics can be checked by test.

It is very important at this stage of development that the design engineer decide definitely how the equipment is to be tested, maintained, and calibrated. Having been in the process of testing the unit himself, he is in an excellent position to determine which test points are most important, which adjustments are critical, and what criteria will be used in making the adjustments. Questions such as these must be answered before the design reaches prototype stage so that proper cognizance can be taken of them in the mechanical layout of the prototype models.

It is likewise important that the design engineer make absolutely sure that the unit will operate properly with the rest of the units of the system. In many instances it is necessary for him to assist in the testing of the other units (which are presumably in the same stage of development) and to invite the design engineers of the other units to test his own. Only by close collaboration at this stage of development can the final product be assured reasonable chances of success. In some instances, it may be necessary to make a combined test of several units in a system before they pass out of the experimental-chassis design stage. This is usually undertaken only if it is more efficient than making a precise measurement on the characteristics of each unit.

9. After the efficiency and reliability of the circuits have been established, prototype model construction can proceed. If the prototype model is to be of any value to the manufacturer, it must contain the same components that the engineer recommends be used in the final product. Hence, it is very important that he check the availability of the components which he recommends and satisfy himself that the unit will pass all of the tests when the recommended components are used. It is also advisable that the designer consult the production engineers who will have charge of producing the unit. This enables him to rectify conflicting views and establish the necessity of following the design principles that he recommends.

10. After the prototype model is constructed, it is subjected to severe testing. This step is the final laboratory proof of the design. If the model responds satisfactorily to these tests, it is considered to be ready for incorporation into the system and tested as a component part of the system.

If the design process has been carefully followed, prototype model tests do not show many unexpected characteristics. It rarely happens, however, that a prototype behaves exactly as expected, and it is usually necessary to make minor revisions in the circuit and in the physical layout. If the necessary revisions are minor, they can usually be accomplished on the prototype model without retracing the earlier design steps and building a new model.

Table 19·3 gives the design process in outline form; the numbered steps correspond to the discussion.

TABLE 19·3.—STEPS IN THE DESIGN PROCESS

1. Block diagram of complete system
2. Block diagram of unit
3. Fill in blocks with old and invented circuits
4. Theoretical analysis of invented circuits
5. Experimental checks of separate circuits
6. Experimental checks of complete circuit
7. Experimental tests to determine:
   a. Power requirements
   b. Component tolerances
   c. Operating stability
   d. Required voltage stability
   e. Heat dissipation
   f. Breakdown into electrical subassemblies
8. Circuit review to:
   a. Check for wasted power
   b. Check for unnecessary tubes
   c. Check proper interconnections to other units
   d. Design power supply
   e. Design subassemblies
   f. Decide on method of testing, maintenance, and calibration
   g. Establish final tentative parts list
9. Prototype model construction
10. Prototype model tests

Table 19·4 is a list of phrases that have meaning to the electronic apparatus designer. Mechanical engineers have frequently used a list of words, such as lubrication, life, vibration, etc., to remind them of design details. The list in Table 19·4 is for the same purpose applied to electronic design.

19·7. Construction Practices for Laboratory Equipment.—The term "laboratory equipment" implies small numbers of equipment fabricated from readily available or handmade parts, used and maintained by technically trained personnel, and operated under reasonably constant temperature and humidity conditions. It also implies that weight and size are of secondary importance. Such equipment is characterized by a multiplicity of adjustments, selected components, open construction, and very high performance.

Equipment for military or commercial use implies large numbers of equipment, utilizing special parts and tooling, used and maintained by average personnel under widely variable conditions. Properly designed

TABLE 19·4.—DESIGN FACTOR CHECK LIST

Drift of characteristics with time
Voltage and current ratings
Derating factors
Voltage breakdown
Efficiency
Power output
Power in; VA in
Frequency, voltage limits of sources
Component tolerances
Tolerances of coefficients
Effect of supply voltage changes
Regulation
Hysteresis in voltage, flux, etc.
Effect of temperature
Power factor; Q
Shielding
Effect of earth's magnetic field
Magnetic coupling of transformers
Distributed capacitance, inductance
Lead inductance
Electrostatic pickup
Mutual inductance pickup
Rise above ambient
Ambient limits
Radiation, generation of noise
Input, output impedance
Contact, thermal emf's
Grid current

Heater-cathode voltage
Heater-cathode leakage
Effect of power source subharmonics
Effect of power source transients
Ground currents
Leakage
Condenser "soaking"
Source impedance
Stray capacitance
Impedance-frequency characteristics
Amplitude-frequency characteristics
Phase-frequency characteristics
Nonlinearity, distortion
Voltage-current characteristics
Loading of source
Phase shifts
Photosensitivity
Deionization time
Stability, phase, gain margins
Pressure, density of air
Finishes
Weight
Size, shape, form factors
Cost
Availability in desired tolerances
Mounting means
Packaging for shipment
Difference between manufacturers

Resistance to thermal shock
Life, shelf life
Humidity, salt spray, immersion
Mechanical strength
Inflammability
Electrolysis; corrosion
Vibration, shock, shockmounting
Mechanical resonance
Operating position
Lubrication, wear
Sparking
Fungus; dust
Torque, speed
Smoothness
Nonoperating service conditions
Producible by a specific mfgr.
Generation corrosive or toxic fumes
Use of critical or expensive materials
Adequate name plates
Minimum tube types
Size, insulation of hook-up wire
Overload protection, fusing
Safety in maintenance, operation
Ease of maintenance, operation
Handles
Illumination of dials
Sharp corners
Test points

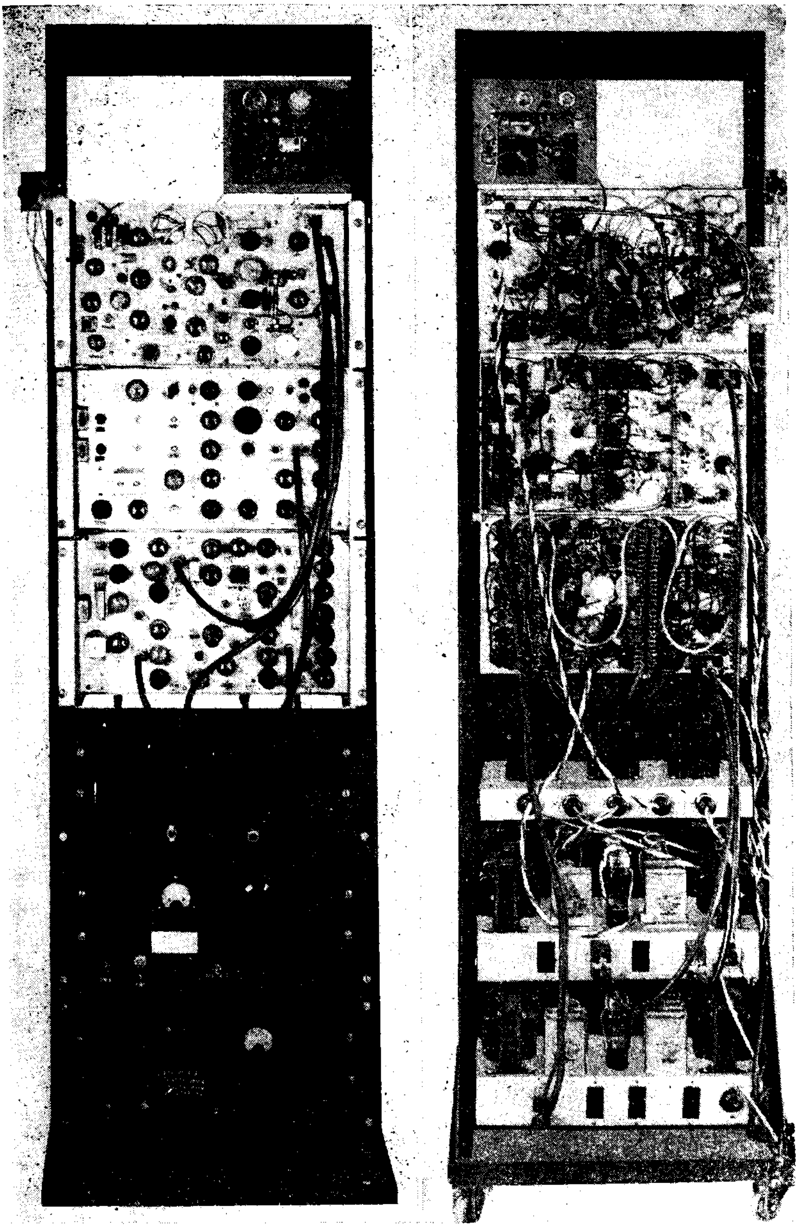


FIG. 19·6.—Example of construction of laboratory equipment using bathtub rack panel.

and constructed equipments of this nature are characterized by few adjustments, standard components, compact construction, ruggedness, portability, and guaranteed minimum performance.

Obviously, these differences between laboratory and commercial equipments have a profound influence on construction practices. Perhaps the greatest difference stems from the fact that laboratory equip-

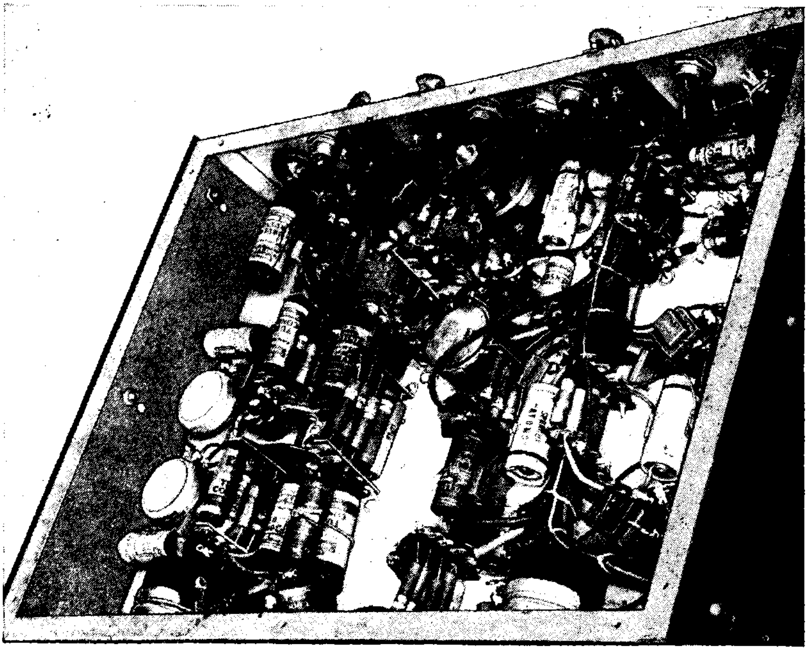


FIG. 19·7.—Example of construction for laboratory equipment with ladder mounting of components.

ment is generally constructed by persons qualified to follow circuit diagrams. This usually means that the only drawing that need be made for laboratory equipment is the circuit diagram. On the other hand, production equipments are constructed by persons who neither see nor have to understand circuit diagrams. At first glance it might seem that what is good practice for one is good practice for the other, but this is not true. If only small numbers of equipments are to be built, it is economically more efficient to hire a more highly skilled workman to build the units from a circuit diagram. If large numbers are to be built, however, it is economically more efficient to spend considerably more time and engineering effort in planning the construction so that it can be carried out by unskilled personnel.

Improvisations are the general rule in laboratory equipment, and the construction is always carried out with flexibility in mind. Laboratory equipment is usually constructed by making maximum use of point-to-point wiring augmented by standard terminal strips. Extra space, unused solder lugs, multiple tie points, etc., are desirable in laboratory equipment. Such "haywire" techniques have no place in commercial production, as they only add confusion in the assembly process.

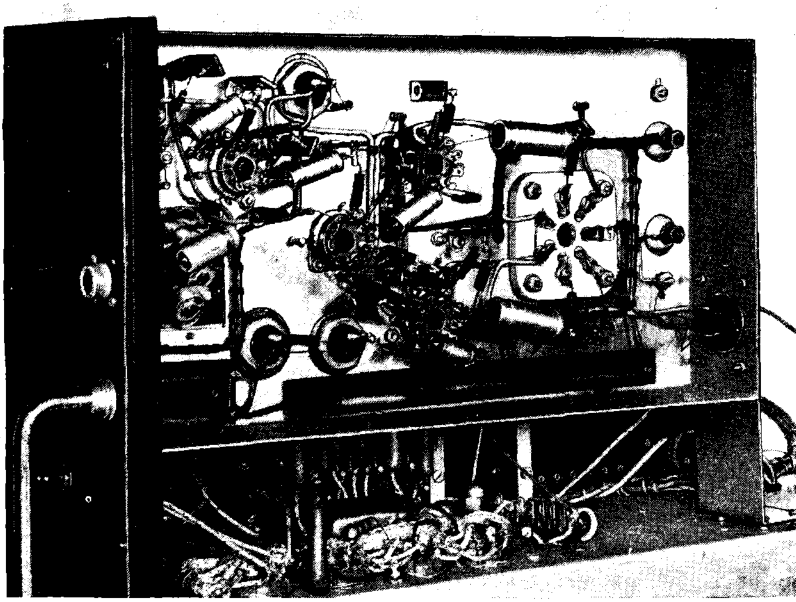


FIG. 19·8.—Example of construction for laboratory equipment.

Many laboratory equipments never get beyond the electrical mockup stage illustrated in Figs. 19·6 and 19·7, which, after all, are perfectly usable forms. Generally, however, laboratory equipment is carried at least one step further, and the circuit is incorporated in a semipermanent fashion on a chassis. Figure 19·8 illustrates laboratory construction.

19·8. Construction Practice for Commercial Equipment.—This section is intended to correlate and augment the various comments that have been made on the design and construction of commercial equipment.[1]

[1] The term "commercial equipment" is intended to cover both industrial and military applications. Construction practices for industrial and military applications differ mainly in the rigidity of the specifications; the general features of design and construction are the same. The term "commercial equipment" does not include equip-

Since electronic equipment is being considered, the fundamental requirement for construction practices is that the electrical circuit will operate satisfactorily. This may or may not be a serious limitation on the method of construction; it depends upon the units under consideration and whether or not the requirements for a satisfactory electrical operation are in conflict with some of the other requirements listed below. For many applications, especially where the circuit contains high frequencies, high voltage, high-gain amplifiers, etc., the physical configuration of the circuit *is* important. In designing the layout it is necessary to consider cabling losses; distributed capacity; distributed inductances; electrostatic, electromagnetic, and microphonic pickup; etc. In many instances, it is impossible to take proper cognizance of all of these variables, and, as a result, prototype models invariably require a certain amount of reworking. During the carrying out of Steps 7 and 8 of the design process, however, it is possible to discover which circuits are critical, and priority of space, location, and attention is given to them in the model construction. Circuit operation can almost always be improved by dividing the unit into separate subassemblies. This does not mean subassembly in the sense that the tubes and heavy components are mounted on a chassis while the light components are mounted on removable terminal strips. What is meant here is the division of the circuit into separate mechanical subunits that when fastened together on a framework form the complete unit. This type of construction has found considerable application in lightweight equipment because it is more economical from the standpoint of space and weight. However, it has a distinct advantage as regards operating reliability as well, which is the point that is stressed here.

The second major consideration is that the mechanical construction be sound. The unit must be mechanically rigid enough to withstand the knocks of normal handling and shipping. It has been found through experience that few units fail to pass vibration tests if they are mechanically rugged enough to withstand ordinary shipment! Adequate mechanical construction applies also to protection of knobs, dials, switch handles, cathode-ray tube faces, and other parts of the unit that are mounted external to the chassis. Locking devices are employed to protect moving parts from accidental movement. Such devices are used to secure large tubes, tuning dials, screwdriver potentiometer adjustments, etc.

One phase of Step 8 in the design process is the determination of the

---

ment for private use, such as home radio receivers. Because of the highly competitive nature of the field of home appliances, the construction practices employed in manufacturing them are necessarily radically different from those employed in manufacturing equipment for military and industrial uses.

best method of testing, maintenance, and calibration. These considerations are vital to the proper operation of any unit; hence, the construction must take them into proper account. The process of testing and calibrating usually requires the measurement of certain voltages, currents, waveforms, etc., and the designer takes proper account of this requirement by marking the critical test points. Proper construction facilitates access to these test points, bringing them to one test panel if possible. Testing and calibration is greatly facilitated if all of the

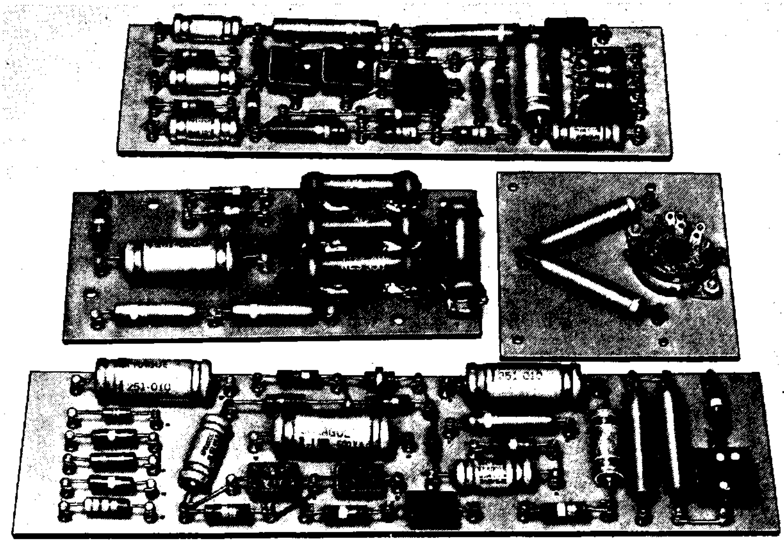


FIG. 19·9.—Typical component strip subassemblies.

test points and all of the calibration controls are located together and are easily accessible.

The quality of maintenance that is performed on any unit is greatly influenced by the ease with which the maintenance can be carried out. The general principles to be observed in constructing a unit for ease of maintenance are as follows: The maintenance should be performed from the operator's position, i.e., the unit should come apart from the front; maintenance must be possible without turning off the unit, disconnecting cables, etc.; maintenance must be possible with the simplest of tools; any tool other than a straight screwdriver and a crescent wrench is considered special by the average maintenance man. The use of snap nuts, cowl fasteners, captive bolts, riveted nuts, wing and thumb nuts, flexible couplings, etc., greatly facilitates access to maintenance and calibration points.

Provision must be made for adequate cooling and protection against moisture, dust, vibration, and shock. Manufacture is facilitated by dividing the construction into a multiplicity of separate operations. Special component strips, illustrated in Fig. 19·9, permit independent fabrication of many subassemblies.[1] Maximum use of mechanical subassemblies is discussed in later sections; however, some mechanical subassemblies are necessary in almost every unit. Figure 19·10 is a 30-megacycle if receiver strip with approximately 20-db gain per stage. It is virtually impossible to obtain satisfactory operation of such units by any other method than point-to-point wiring as illustrated in the figure.

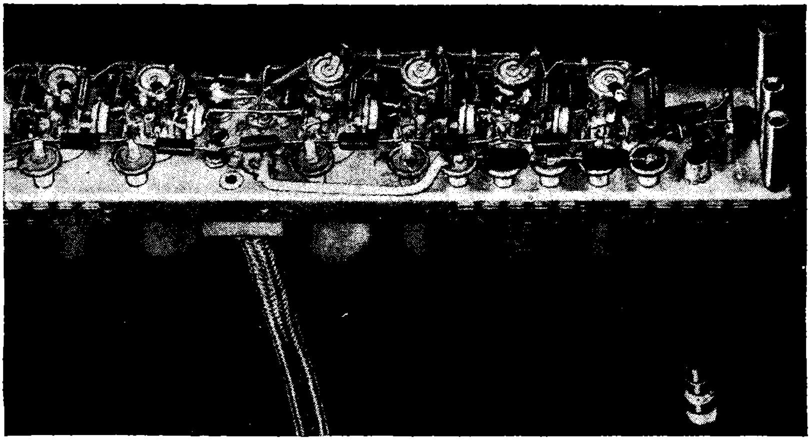


FIG. 19·10.—Radar receiver strip, a typical mechanical subassembly.

Such units require considerable care to provide proper shielding and hence are nearly always constructed on mechanical subassemblies which are in turn mounted to the major units in which the circuit operates.

This section would not be complete without some discussion of the relative merits of open and closed units. Examples of open-unit construction are shown in Figs. 19·11 and 19·12. In operation, of course, these units are covered with thin metal dust covers. In this type of construction each component is either hermetically sealed or constructed to withstand the expected operating conditions. Figure 19·13 illustrates closed-unit construction. In this example the complete unit is enclosed in a pressure tight chamber, and hence the individual components need not be constructed to withstand the full variation of operating conditions.

Closed-unit construction finds its greatest application in airborne

[1] Army Air Force Specification ARL-102-A, Navy Specification RE-13-A-554, and Civil Aeronautics Authority Manual 16 contain much valuable information on the best way of handling component strips, wiring harnesses, tie points, etc.

systems that are expected to operate satisfactorily at very high altitudes. Such construction generally results in considerable saving of weight

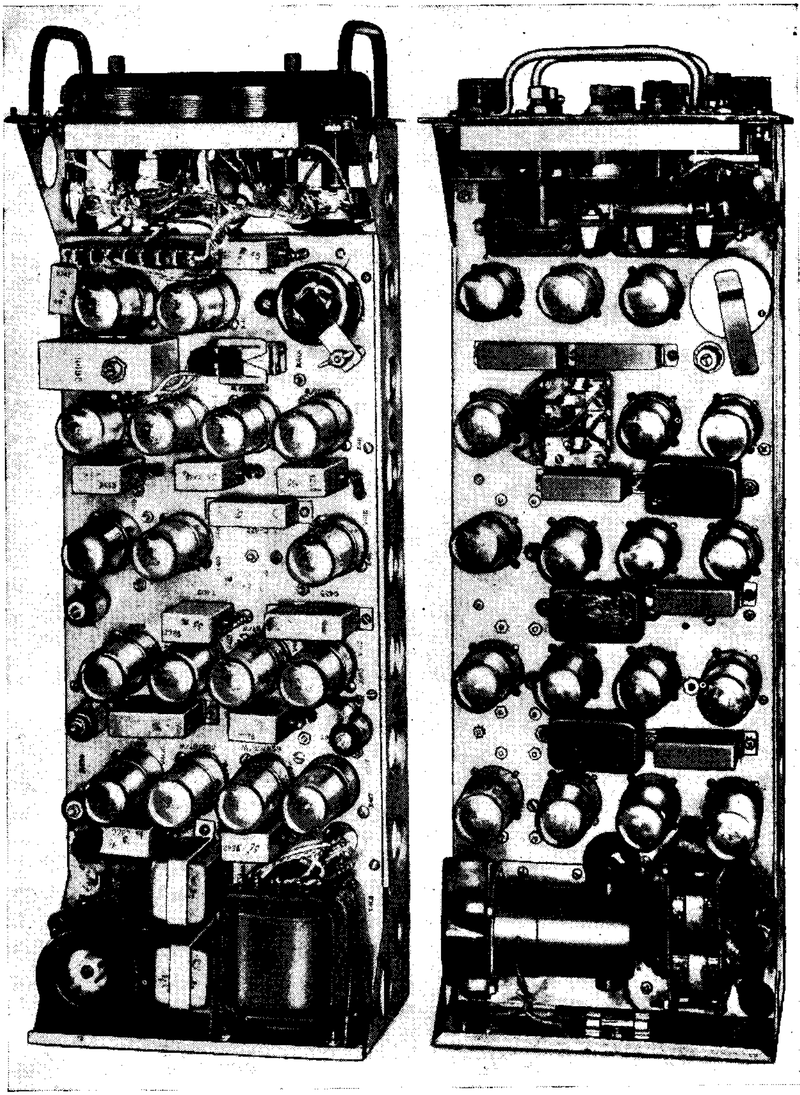


FIG. 19·11.—Bombing radar range units, top view. Example of open construction.

and space, in spite of the added weight of the pressure chamber. However, units that are constructed in this manner are generally more difficult to service because of the necessity for breaking the pressure seal to get to test and calibration points. Problems of heat dissipation become quite severe for closed-unit construction and may require both internal and external blowers.

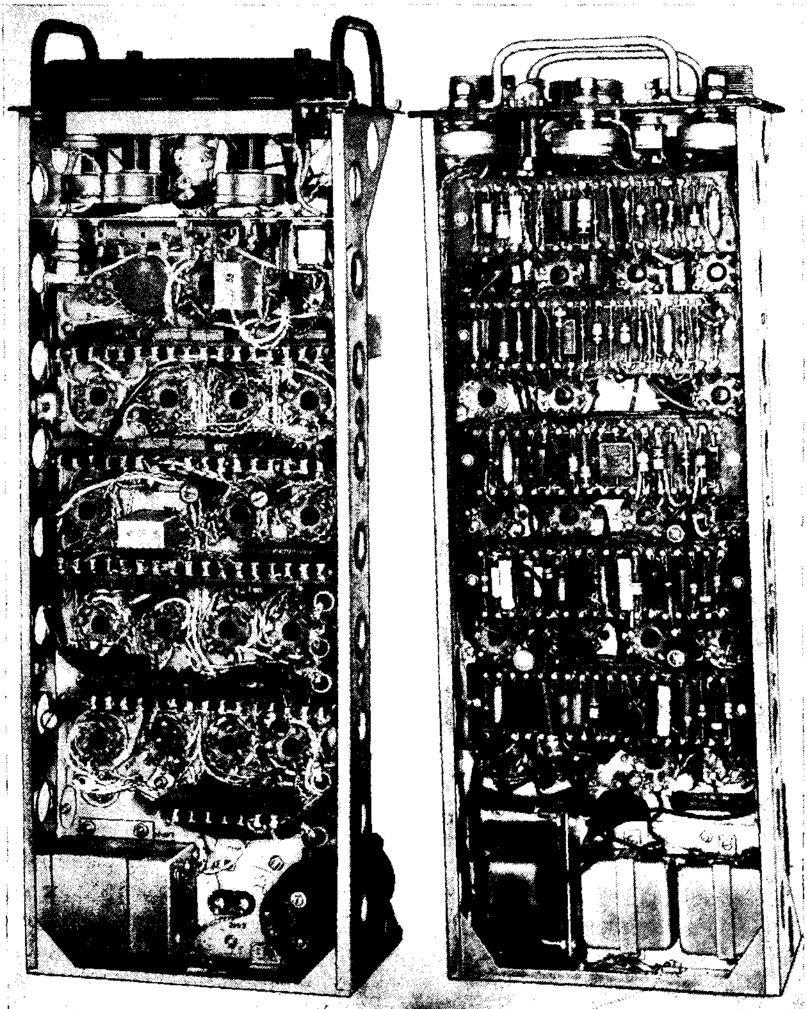


FIG. 19·12.—Bombing radar range units, bottom view. Example of open construction.

Many of the reasons for resorting to closed unit construction are gradually being eliminated. The development of new dielectrics, transformer impregnants, and small hermetically sealed components makes open-unit construction more and more practical in comparison with closed-unit construction. To some extent these improvements in component construction are offset by continuing improvements in aircraft design which enable them to operate at greater altitudes. In all probability, closed-unit construction will continue to be preferred for modulators, r-f generators (in the microwave range), high-voltage sources, and other special units for airborne use.

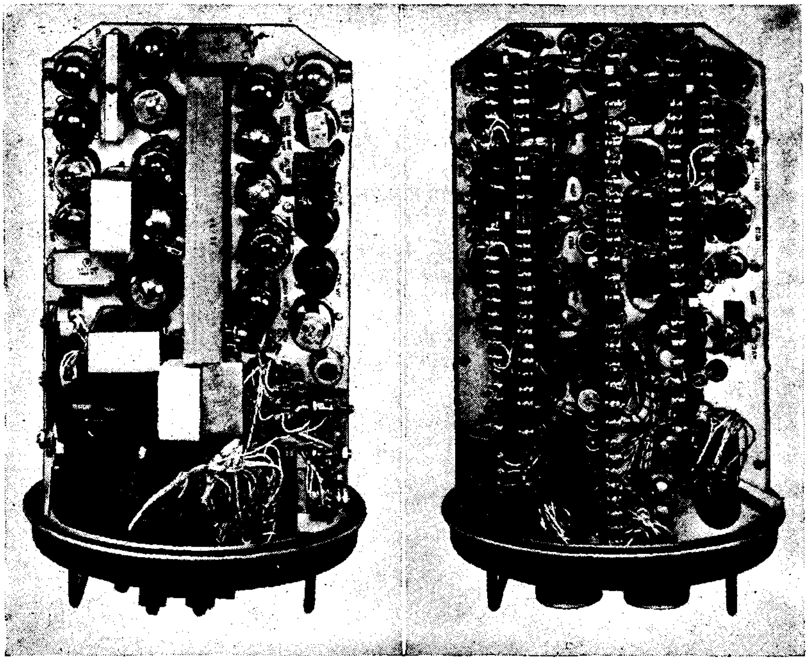


FIG. 19·13.—Example of closed-unit construction. Cover not shown.

19·9. Mechanical Assemblies in Electronic Apparatus.[1]—It is impossible for the designer, whether physicist or electrical engineer, to escape the mechanical aspect in the design and construction of electronic equipment. Mechanical responsibility in design encompasses the practical geometry of the device, the correct choice of materials, finishes, lubrication, cooling facilities, etc. The parts to be fabricated are designed so that they may be made with the available facilities, and a complete set of prints is prepared. Fundamentally, all electronic equipment is mechanical. Since this phase of production is so important, it is advisable for the designer to enlist the aid of a mechanical engineer and assign to him the mechanical aspects of the design process.

[1] Sections and 19·9 and 19·00 are by A. C. Hughes, Jr., and J. V. Holdam, Jr.

Mechanical assemblies are divided into several classes on the basis of the type of engineering effort necessary to produce each. Components, such as resistors, condensers, etc., are mechanical assemblies; their mechanical characteristics are given in the component specification. The machinery for making components is another class. The development of such machinery may be necessary if the design employs a new or specially built component. Housings, holders, brackets, connectors, tube sockets and clamps, chassis decks, front panels, terminal strips, etc., form a third class and are generally available from commercial manu-

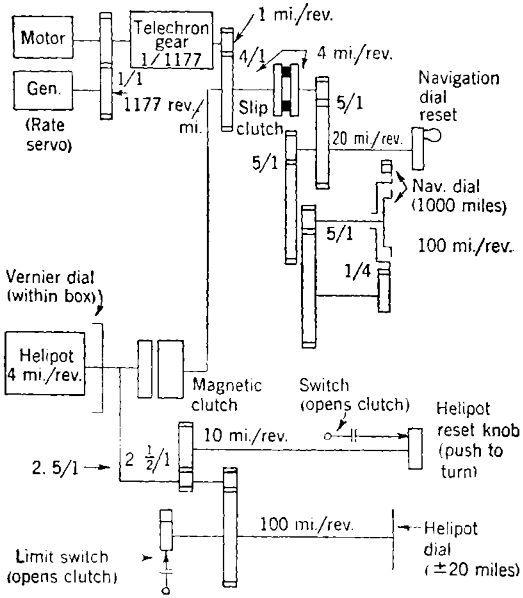


FIG. 19·14.—Mechanical schematic diagram.

facturers in an adequate mechanical design. However, when space and weight are critical, it is usually necessary to make special designs of some items in this class. The fourth and most important class is assemblies with moving parts. This class is further divided into electrical and mechanical assemblies. For example, potentiometers, variable resistors and condensers, motors and generators, relays, magnetic clutches, synchros, and switches are electrical; dial and counter mechanisms, hand-operated gear trains, servomechanism gear trains, mechanical computer elements, and drive mechanisms are mechanical.

The design of mechanical assemblies is carried out in the same sequence as the design of the electrical circuits. The first six steps given in the design process in Sec. 19·6 apply almost directly to mechanical design;

Steps 7 and 8 may apply only partially; and Steps 9 and 10 apply directly. In general, the mechanical design lags the electrical design because of the greater time required to try out new ideas. Also, the expense of making mistakes in the mechanical design is greater because of the shop time that is wasted. When the mechanical characteristics cannot be synthesized, the early stages of electrical design include some mechanical assemblies; but they are fabricated in very crude style. The counterpart of the electrical mockups used in the early electrical design is the mechanical breadboard; a part of Vol. 17 of this series is devoted to the use of mechanical breadboards.

The successful completion of the first six steps in the design process supplies sufficient information on the mechanical requirements of the unit to finish the mechanical design. If the mechanical assemblies are at all complicated, it is advisable to make a mechanical schematic drawing on which are specified the various electrical and mechanical inputs, the gear ratios, shaft speeds, etc. Figure 19·14 is an example of this kind of schematic diagram.

The construction details are determined by consideration of the power and torque requirements of the assembly and the conditions under which the equipment must work, etc. This phase of the design is straight mechanical engineering and is adequately covered by current engineering literature.

19·10. Lightweight Construction.—The general circuit design considerations that require emphasis when the design and construction of a unit is to result in minimum weight and space have been partially covered in Sec. 19·9. This section deals with the problem from the standpoint of construction and points out the steps in the design process (Sec. 19·6) where an effort toward lightweight design and construction is most rewarding.

Presumably, the desire for lightweight construction is known from the outset, and a great deal of progress is made in Steps 1 and 2 of the design process. During the system and unit block diagram stages, it is advantageous to consider the pros and cons of unit size, cabling, power supply, and power dissipation.

Equipment serviceability, a very important characteristic, gets progressively poorer as the units are combined into larger units. Also, large units cannot be so advantageously located as small ones, since they require special mounting provisions. This often tends to counter-balance the reduced inter-unit cabling by requiring even longer cable runs between the units that are used.

For a general case these arguments can be summed up as follows. The total weight and volume of a system become progressively smaller as the units are combined into larger and less numerous units. The serviceability and flexibility of location tend to become poorer as the units are combined into larger and less numerous units. The compromise that is reached in any particular system depends on the system itself and on the application for which it is designed. There is usually a best compromise for each system, however, and reaching it requires a thorough analysis of the variables.

The next phases in the design process that require particular emphasis in lightweight design are Steps 7, 8, and 9. During the experimental tests and general circuit review, particular emphasis is placed on reducing the total amount of power that is consumed by the unit. Operation of the tubes at the minimum possible plate voltage and the selection of tubes for low heater dissipation have already been mentioned. There

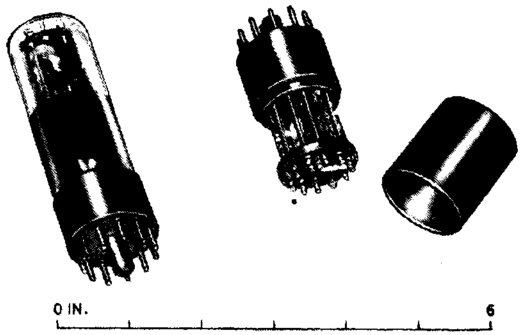


FIG. 19·15.—Delay multivibrator subassembly.

are at least two other methods by which the total power consumption may be reduced. In amplifier circuits where it is necessary to maintain a definite pass band, the value of the plate resistors is usually determined by stray capacity of the load and the shunt capacity of the tubes. By employing construction techniques that minimize unwanted capacity and by selecting amplifier tubes for minimum shunt capacity, the pass band can be maintained even though the plate resistors are made quite large. This materially reduces the quiescent power dissipation in an amplifier circuit. Tubes that operate on time asymmetrical waveforms can often be arranged in the circuit so that the tube is cut off most of the time.

It was pointed out in the discussion of the design process (Sec. 19·1) that efficient power-supply design requires careful consideration of multi-bus supplies, the use of available direct current for the tube heaters, and a careful analysis of the minimum regulation required.

Subassembly Design.—Experience at the Radiation Laboratory has shown that a very profitable approach to lightweight design is careful use of the subassembly technique. The approach to this problem has been made in two rather different ways, both based on the use of miniature and subminiature tubes and very small components. Neither of these techniques is much beyond the development stage, but prototype models employing them have been built and show great promise in reducing the weight and space of most units.

Figures 19·15 and 19·16 are illustrations of the "circuit-within-a-tube" approach to subassembly design. The advantages of this type of construction are many fold. The unit can easily be hermetically sealed; replacement of the circuit is very simple; and the performance of the circuit can be held to close tolerance. Hermetic sealing is advantageous,

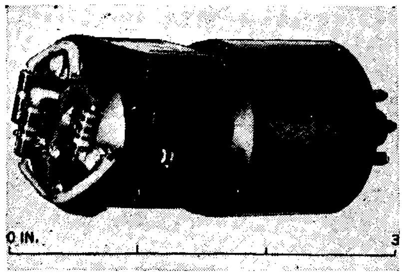


FIG. 19·16.—Twin-T subassembly.

since the individual components are not required to operate over the wide temperature and humidity range to which the complete unit is subjected. By filling the envelope with a dry inert gas such as helium, heat dissipation is facilitated, leakage is reduced to the minimum, and the possibility of voltage breakdown when the unit is operated at high altitudes is minimized. Such a unit is changed with the same ease that an ordinary tube is changed. Hence this type of construction increases the serviceability of the unit. By adjusting the values of the components to match the characteristics of the tube, during the subassembly fabrication, the performance of the circuit can be held to much closer tolerances than if this type of construction were not employed. The price, of course, for this advantage is the increased cost of the subassembly; no attempt is made by the repairman to service the subassembly itself.

The unit shown in Figure 19·16 is a twin-T filter. This unit was produced in very large quantities for use in an amplifier of special design that was part of an airborne radar system. The bandpass and attenuation characteristics of the twin-T filter were very critical, and it would have been impossible to guarantee satisfactory operation of the unit in which the twin-T was employed had this type of construction not been used. By dividing the series and shunt resistances into pairs, nonprecision parts were employed, the padding resistor being selected to compensate for the tolerance in the associated resistor. This is an illustration of application of this type of construction to obtain performance within very close limits while using components that are not held within close limits.

Another type of construction that holds great promise by reducing the unit size and weight is illustrated in Figs. 19·17 to 19·19. The unit shown is an experimental airborne navigational range unit. It contains 50 tubes, weighs 18 lb, and is approximately 12 by 10 by 8 in. The details of the subassembly construction are shown in the figures. This

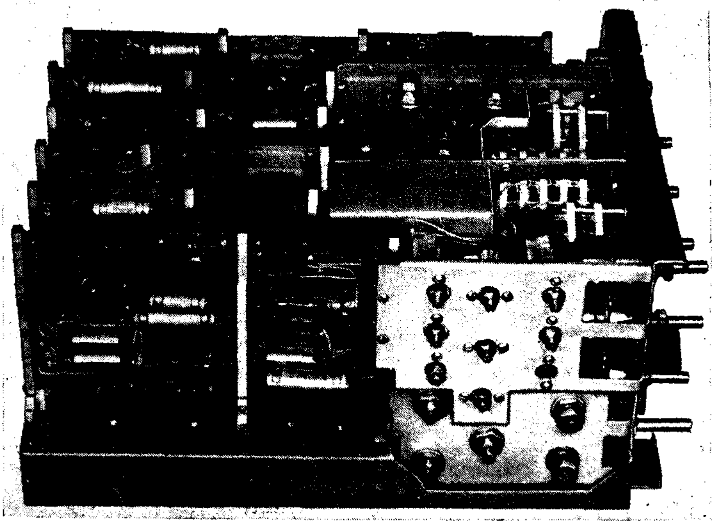


FIG. 19·17.—Airborne navigational range unit using subassembly cards.

type of subassembly has most of the advantages, except hermetic sealing, of the subassembly methods discussed above; it has the additional advantage of being considerably more flexible, since larger circuits can be built on one subassembly. By building a complete functional circuit on one subassembly, stray capacity is reduced to the very minimum; hence maximum power efficiency is obtained by use of large plate resistors. In the unit that is taken as an example of this type of construction, the subassemblies are held by grooved support guides; the dust cover keeps the subassembly plug mated with the chassis socket. The principal adjustments are either on the front panel or on one side, easily accessible for maintenance or operation. Since all of the intercircuit wiring passes through the connectors underneath the chassis, this type of construction is admirably suited for the inclusion of many test points. Figure 19·20 is a block diagram of the complete unit Each of the various blocks is a separate functional circuit. Figures 19·21 and 19·22 show the circuit and layout of one of the subassemblies.

The unit is made very compact by holding the tubes with clips rather than tube sockets and by mounting all of the "large" components on one side of the terminal strip and small components on the other side. This permits the terminal strips to be stacked very closely together on the chassis.

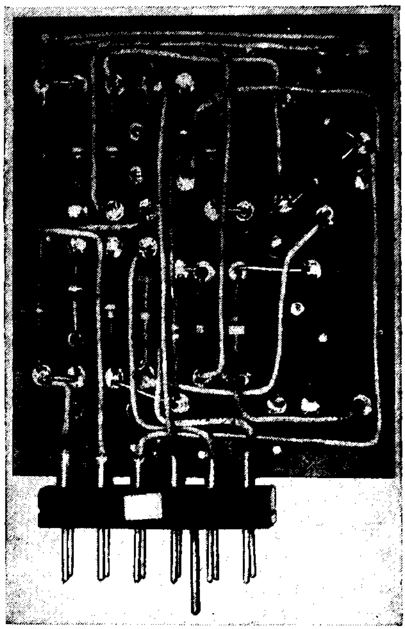


FIG. 19·18.—Details of subassembly construction.

The principal disadvantage of this type of construction is the large number of connectors that are used. Completely suitable connectors are not commercially available. However, this type of construction is very promising, and it appears to be fundamentally sound. It is admirably suited to both production and maintenance, since the subassembly construction permits multiple independent fabrication and very easy replacement of malfunctioning units.

19·11. Lightweight Mechanical Assemblies.[1]—The following discussion is concerned with lightweight mechanical assemblies used in radar, radio, and other electronics equipment. In many cases well over half of the weight of a radar set is taken up by mechanical assemblies such

[1] Section 19·11 is by A. C. Hughes, Jr.

as revolving scanners, gear trains for scope indicators, servo gear trains in general, and azimuth and range dials, counters, etc.

One of the simplest rules in designing lightweight equipment is to specify light materials such as alloys of magnesium and aluminum. These lighter metals cannot replace steel and brass completely, since shafting, bearings, some gears, cams, and general wearing surfaces may require the heavier metals; but framework, castings, brackets, covers,

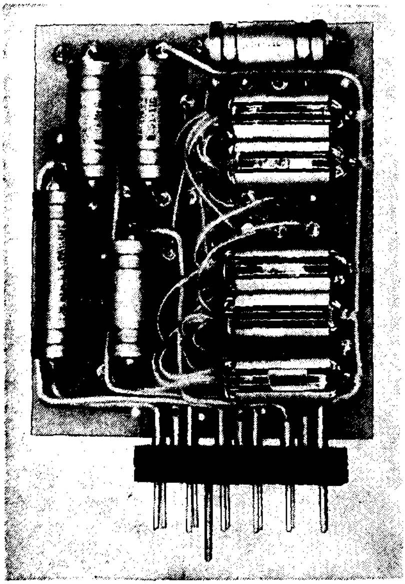


FIG. 19·19.—Details of subassembly construction.

spacers, handles, cranks, some gears, chassis decks, etc., can be made from lighter metals with appreciable savings in weight.

There are two well-known factors that require careful attention when light metals are used in conjunction with heavy metals—the difference in coefficient of linear expansion and electrolysis.

Where there are many parts and the design is complicated, wooden mockups or cardboard models are valuable. Seeing the design in three-dimensional solid pieces is very helpful in making the assembly compact

Figure 19·23 shows an example of mechanical design of rather high space-use. The device shown consists of a constant-speed motor, speed-reducing gear unit, a slip clutch, and two potentiometers. Between the two potentiometers are an end-return torsion spring, a mechanical-end stop, and a cam and push rod to operate a limit switch. The unit weighs 32 oz and is 2 by 4 by $5\frac{1}{4}$ in. in size.

As the density rises, maintenance and replacement become more difficult. Subassembly methods of construction tend to alleviate those

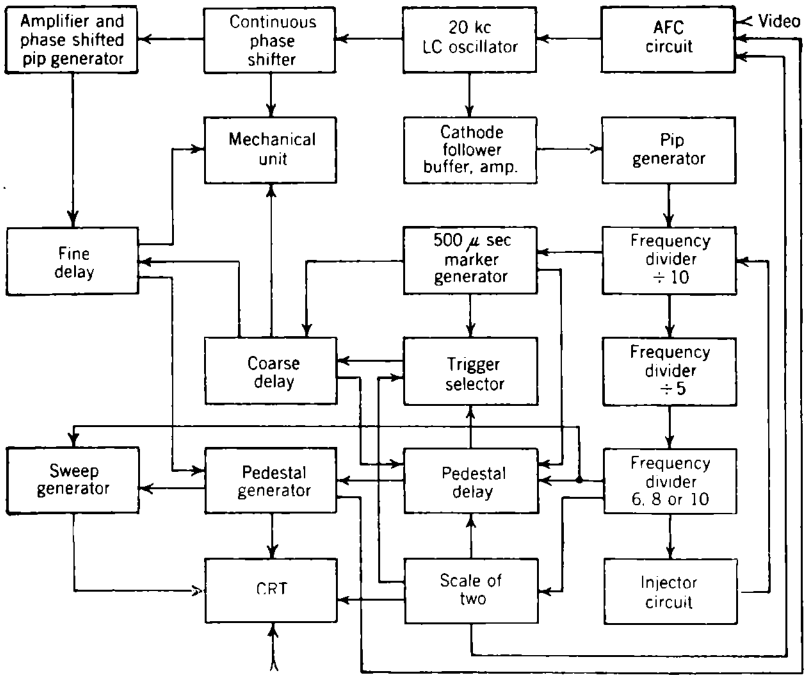
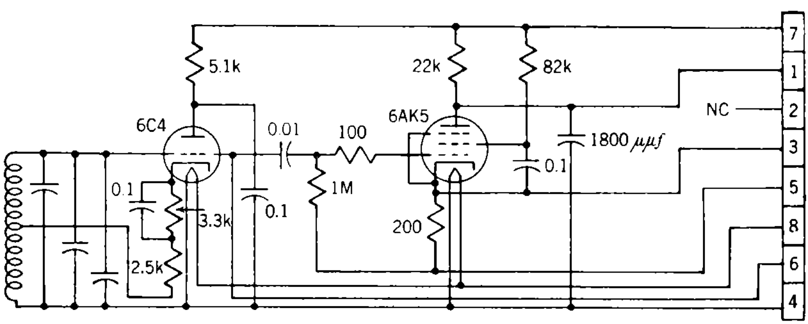


FIG. 19·20.—Block diagram of airborne navigational range unit.



Sub-assembly construction

FIG. 19·21.—Circuit of 20-kc oscillator subassembly.

difficulties for mechanical assemblies as well as for electrical assemblies. A mechanical subassembly is usually thought of as a group of parts having no permanent mechanical connection to any other mechanical parts or group of parts. Since mechanical assemblies in electronic equipment are usually closely related to one circuit, usually an electrical subassembly in itself, it is often convenient to make the mechanical assembly and the circuit one subassembly. Figure 19·24 shows an

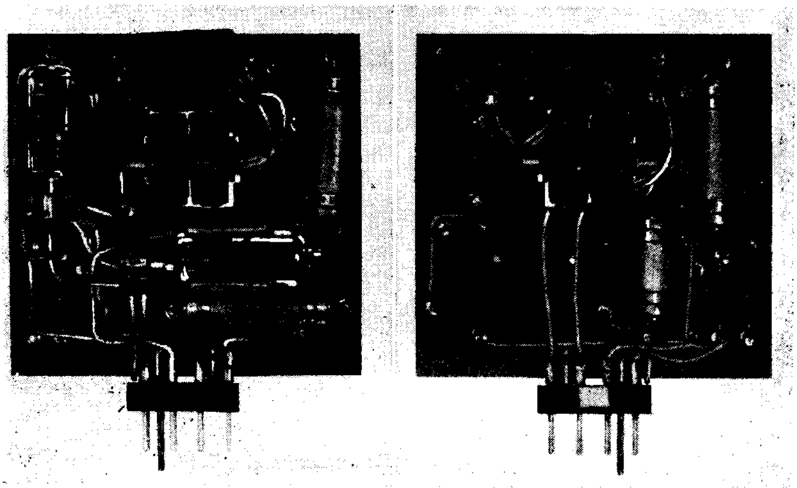


Fig. 19·22.—Subassembly of 20-kc oscillator.

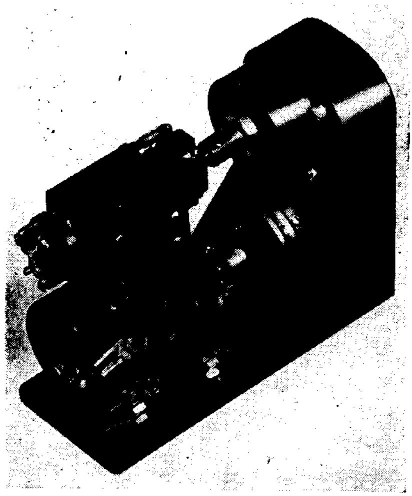


Fig. 19·23.—Lightweight mechanical assembly.

experimental mechancial subassembly and electrical subassembly mounting card joined in this manner. The saving in space, wire, and connectors is appreciable. Electrical parts can be fitted into voids and spaces of the mechanical unit that would otherwise not be filled. Connector problems and the weight of connectors may be avoided by wiring all the parts of a mechanical unit directly into the electrical unit of the same subassembly. Where mechanical units are subassemblies by themselves, it is an advantage to have them automatically plug in as they are fastened into place.

19·12. Summary.—The importance and use of specifications are discussed. The fundamentals of temperature-rise analysis are reviewed and equations for the cooling of open and closed apparatus by convection and forced circulation are included. A systematic design procedure and a check list of design factors are described. Frequent reference to this list is made in other chapters. The design and construction of lightweight,

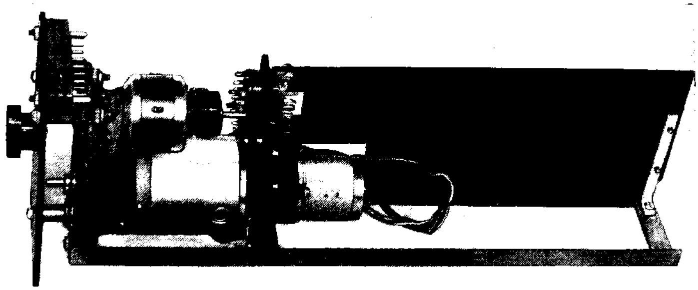


FIG. 19·24.—Mechanical and electrical subassembly.

minimum heat-dissipation electronic apparatus is treated. Important points covered include the following: the use of both mechanical and electrical subassemblies; reduction of heater dissipation; reduction of plate voltages; normally-off operation of tubes; reduction of voltage-divider dissipation; provision of separate regulated screen voltages; use of the minimum amount of voltage-supply regulation; and elimination of excess stray capacitance, thereby allowing load impedances to be increased without loss of bandwidth. Design and construction differences between laboratory and commercial or military apparatus are briefly considered.

Index

S