RCA REVIEW

a technical journal

WHC

RADIO AND ELECTRONICS
RESEARCH • ENGINEERING

VOLUME VIII SEPTEMBER 1947 NO. 3

RCA REVIEW

George M. K. Baker
Manager

Chas. C. Foster, Jr.
Business Manager

SUBSCRIPTIONS:

United States, Canada and Postal Union: One Year $2.00, Two Years $3.50, Three Years $4.50
Single Copies: 75¢ each

Other Countries: One Year $2.40, Two Years $4.30, Three Years $5.70
Single Copies: 85¢ each

Published quarterly in March, June, September, and December by Radio Corporation of America, RCA Laboratories Division, 30 Rockefeller Plaza, New York 20, N. Y.

Editorial and General Offices: RCA REVIEW, Radio Corporation of America, RCA Laboratories Division, Princeton, New Jersey.

Entered as second class matter April 3, 1946, at the Post Office at New York, New York, under the act of March 3, 1879

RADIO CORPORATION OF AMERICA
David Sarnoff, *President*

Lewis MacConnach, *Secretary*
Arthur B. Tuttle, *Treasurer*

PRINTED IN U.S.A.

RCA REVIEW

a technical journal

RADIO AND ELECTRONICS
RESEARCH • ENGINEERING

Published quarterly by
RADIO CORPORATION OF AMERICA
RCA LABORATORIES DIVISION
in cooperation with

RCA VICTOR DIVISION
RADIOMARINE CORPORATION OF AMERICA
RCA INTERNATIONAL DIVISION
RCA COMMUNICATIONS, INC.
NATIONAL BROADCASTING COMPANY, INC.
RCA INSTITUTES, INC.

VOLUME VIII SEPTEMBER 1947 NUMBER 3

CONTENTS

RCA REVIEW

IN MEMORIAM
JAMES G. HARBORD

It is with deep regret that *RCA REVIEW* announces the death of Lieutenant General James G. Harbord, USA (Retired), former President and Chairman of the Board of the Radio Corporation of America. General Harbord's death at the age of 81, came after a brief illness at his home, Dogwood Lane, Rye, N. Y. on August 20, 1947. At that time he was Honorary Chairman of the Board of RCA.

General Harbord's distinguished military career needs no detailed presentation here. From the time of his entering the Army as an enlisted man in 1889, through the swift-moving events of his rapid rise to appointment as General Pershing's Chief of Staff of the First AEF and Chief of the Services of Supply in World War I, his subsequent duties as Deputy Chief of Staff of the U. S. Army following the war, to his retirement in 1922, General Harbord displayed the qualities of leadership and courage which prompted the following words from the Secretary of War:

> "The industry into which you are going is still in its infancy, and offers a large field for your activity. Being a gallant leader in the Army, there is no doubt that you will prove to be an equally great leader in the industrial and commercial field."

That these words were well chosen and thoroughly applicable to General Harbord has been proved many times in the last two and a half decades. In 1922, he was elected a Director and President of the Radio Corporation of America. In 1930 he became Chairman of the Board of Directors in which capacity he served until his retirement shortly before his death. At that time General Harbord was also serving as a Director in several other corporations.

General Harbord gave active support and encouragement to *RCA REVIEW* since it was first published in 1936. His interest and friendly counsel contributed in no small measure to the success of this journal.

FOREWORD

AN **Engineering Book Series** has recently been added to the publishing activities of the RCA REVIEW Department of RCA Laboratories Division. The first volume in this new book series is entitled PATENT NOTES FOR ENGINEERS.

RCA REVIEW publications now include:

(a) *RCA REVIEW* — technical journal containing research and engineering papers written by RCA authors covering radio, electronics and related fields. ***Published quarterly.***

(b) *RCA Technical Papers—Index*—chronological and alphabetical listings of substantially all papers written by RCA authors. Indexes are cross-referenced by subjects and authors. ***Published yearly***, with quarterly listings in each issue of *RCA REVIEW*.
Currently available: RCA TECHNICAL PAPERS—INDEX, *Volume* I (1919-1945)
RCA TECHNICAL PAPERS—INDEX, *Volume* II(a) (1946)

(c) *RCA Technical Book Series*—collections of technical papers, written by RCA authors and reprinted from leading journals, on subjects of **major** importance in the radio and electronics fields. A few previously unpublished papers are frequently included in each volume. ***Published aperiodically.***
Currently available: TELEVISION, *Volume* III (1938-1941)
TELEVISION, *Volume* IV (1942-1946)
RADIO FACSIMILE, *Volume* I (1938)

(d) *RCA Engineering Book Series*—original material by RCA authors of a less technical nature than that included in (c), which is nevertheless of interest and value to engineers because of a close relation to their work. Collections of previously published material are also included in this series whenever applicable. ***Published aperiodically.***
Currently available: PATENT NOTES FOR ENGINEERS

(e) *RCA Technical Pamphlets*—miscellaneous material of general interest and value to engineers, including bibliographies, collections of abstracts, notes, etc. ***Published aperiodically.***
Currently available: TELEVISION Bibliography—(1929-1946)

Further information on any of these publications may be obtained by writing to:

RCA REVIEW
Radio Corporation of America
RCA Laboratories Division
Princeton, New Jersey

The Manager, RCA REVIEW

Corrected to: September 1947.

TAPE RELAY SYSTEM FOR RADIOTELEGRAPH OPERATION*

BY

SIDNEY SPARKS† AND ROBERT G. KREER‡

RCA Communications, Inc.
New York. N. Y.

Summary—*In the Tape Relay System described herein, telegrams are received in terminal radio offices over wire and radio channels in the form of printed reperforated tapes suitable for immediate retransmission. Manual letter-by-letter reprocessing both at terminal radio offices and at intermediate forwarding offices is eliminated. Speed of service is improved, possibilities for human errors are minimized, and unit operating costs are reduced by the new methods. The tape relay method of operation, the means for coordinating it with other systems, and the present extent of the RCA Tape Relay Network are discussed in this paper.*

INTRODUCTION

In order to meet the growing and more exacting demands of the public for faster, more accurate and more dependable radiotelegraph service in the international field, there has been developed and made effective the plan of operation described in this paper.

RECENT HISTORY

Prior to World War II most radiotelegraph circuits throughout the world were operated by the Morse method, and had been so operated without basic change for some twenty years. In the early stages of World War II, under the pressure of unprecedented demand for radiotelegraph service, the United States Army pioneered in the use of apparatus and procedures in its vast networks which substantially changed the outlook of the radiotelegraph industry.

Basically the Army system was built around improvement of radio circuits to the fullest practicable extent, including use of directional antennas, wise choice of operating frequencies, and use of frequency shift transmission, so that printing telegraph tape relay methods could be more effectively employed. Use of the tape relay method, which had been developed to a high measure of efficiency in land-line operation, permitted traffic to flow freely between land-line and radio links with-

---

* Decimal Classification: R530.
† Vice President and Traffic Manager. ‡ Traffic Engineer.

out the necessity of manual letter-by-letter reprocessing at relay points. This permitted fast service with minimum risk of errors and at low unit cost.

Engineers and management personnel of this company were engaged during the war in the creation and operation of the Army networks, and were intimately familiar with the methods used and with the operating results. At the conclusion of the war, elements of the military plan adaptable for commercial purposes were combined with other new features to create a system of printing telegraph tape relay operation which has become familiar to communications people throughout the world.

The most important military networks were operated by simple and efficient tape relay methods based upon the five unit printing telegraph code for which an extensive line of apparatus had been developed through years of experience in the land-line systems of the United States. The plan of operation in use and described herein also is based upon these methods to the fullest practicable extent, but cognizance is taken of the practical impossibility of quickly converting all commercial radio circuits to the newer method of operation. For an interim period of indeterminate duration the plan contemplates the use of code converters for automatically translating Morse and other codes into five-unit code and vice versa. Thus, at least some of the advantages of tape relay operation can be gained even while older methods of operation must be used on some radiotelegraph channels.

Printing Telegraph Tape Relay Operation

Advantages of Five Unit Code Tape Relay System

The aim of tape relay operation is to achieve maximum speed of service with minimum risk of errors and at low unit cost. This is accomplished by eliminating letter-by-letter manual processing except at the point where a message is prepared for original transmission. Messages are handled through relay points in a tape relay network by simple physical transfer of message tapes.

In older methods of operation, the average message was processed letter-by-letter several times between acceptance from the originator and delivery to the final addressee. Prior to inception of the five unit code tape relay method each incoming or outgoing message was processed into page form as it was received in the central office and then manually reperforated, or retransmitted directly from a keyboard operated transmitter for land-line distribution or for onward transmission over a radio channel. In many cases messages were subjected to the same laborious reprocessing in one or more intermediate central relay offices through which they passed in transit. Each such reprocessing introduced added cost, delay and opportunity for human errors.

Basic Apparatus Used in The Tape Relay System

The Typing Reperforator—The primary unit in this Tape Relay System is the typing reperforator. This instrument, without base or with receive-only base, is used for the reception of messages over wire or

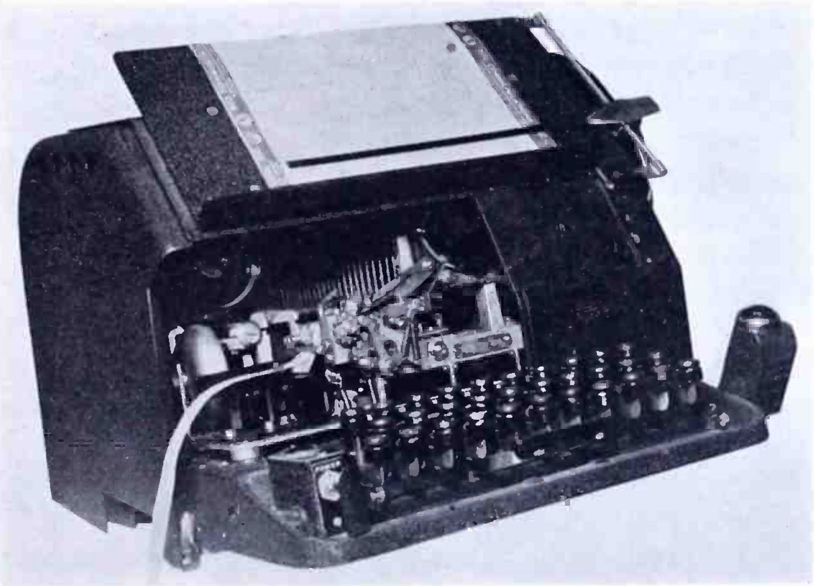


Fig. 1—Send-receive typing reperforator with cover modified to facilitate observation and removal of tapes.

radio channels. In this form it is called a Receive-Only Typing Reperforator. Equipped with keyboard, back-spacer, and end of line indicator, it is called a Send-Receive Typing Reperforator and is used as a printing perforator for the preparation of messages for original transmission in branch and central offices. In each case, messages emerge from the typing reperforator in the form of perforated tapes with the corresponding intelligence printed on the same tape. Use of typing reperforators increases speed of message handling, minimizes risk of errors, by rendering tapes readable at a glance, and contributes to lower operating cost. A send-receive typing reperforator is shown in Figure 1.

The sample tape shown in Figure 2 is of the form produced by the typing reperforator. It is known as chadless tape because the small discs, called chads, which are perforated to form the code combinations are not cut completely from the tape but are perforated only suffi-

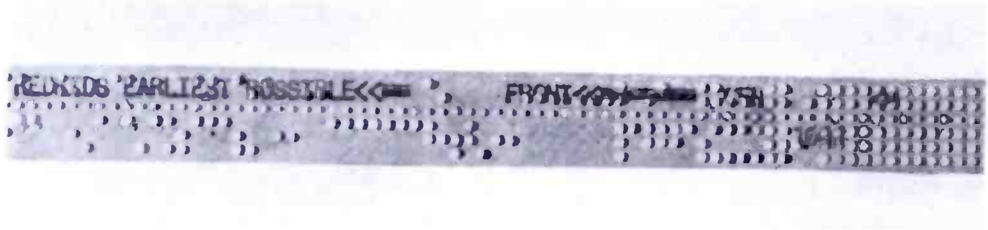


Fig. 2—Five unit chadless typed reperforated tape.

ciently to permit the chads to rise like small hinged lids in response to the sensing pins of a transmitter. The printed characters on this type of tape are six spaces to the right of or behind the corresponding code perforations, and because no part of the tape is actually punched out, they are readily legible. The printing of upper and lower case characters in separate alignment facilitates the checking of tape for channel numbers and proper functional characters. A secondary advantage of the chadless type of perforation is the elimination of the refuse problem in collecting and disposing of the perforated chads.

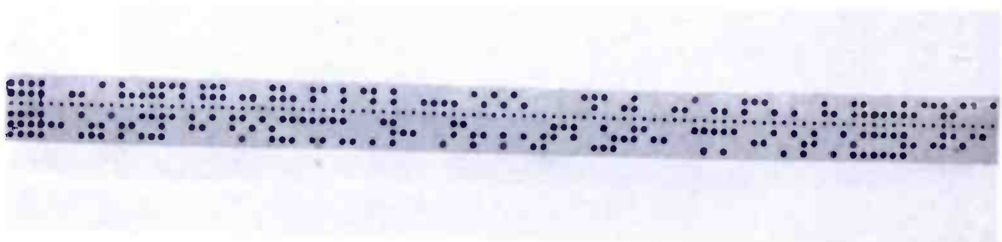


Fig. 3—Five unit fully perforated tape.

Non-typing perforators and reperforators produce fully perforated tapes of the type shown in Figure 3. Particularly in the case of the manually operated keyboard perforator, such devices have the advantage of greater speed of operation when in the hands of highly skilled operators. They are not suitable, however, for general use in connection with the tape relay method due to the difficulty of reading unprinted tapes for the purpose of routing them through relay points and of detecting errors or mutilations.

It may be seen from Figure 4, which shows the entire five unit code,

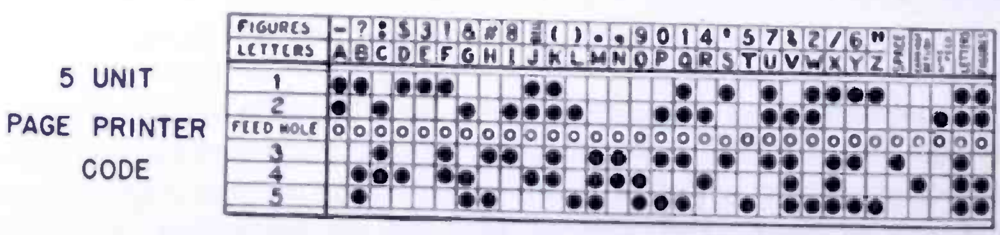


Fig. 4—Five unit printer code.

that all characters are represented in this code by different combinations of five signal elements, each of which may be marking or spacing in nature. The perforations in the tape cause marking signals to be transmitted and unperforated sections represent spacing signals. In the start-stop system, with which this paper is principally concerned, these elements are transmitted between the self-synchronizing start and stop pulses before and after each character. For example, the letter "A" is represented in the tape by two perforations above the small feed hole and a solid area below. This would cause the transmitter to send two marking signals followed by three spacing signals. The normal transmitting speed is 368 operations or characters, nominally sixty words, per minute on radio and wire line channels alike. This adds to the flexibility of the system by rendering apparatus components generally interchangeable in either service. Higher operating speeds are used in appropriate circumstances.

Tape Transmitter Distributors—Advantageous use of automatically reperforated tape is dependent upon suitable devices for retransmitting intelligence perforated into the tape. Such devices are available in several forms. The tape transmitters used in this Tape Relay System are of three types, all of which perform the identical basic function of reading, through the action of sensing pins, the perforations in the tape and transmitting the corresponding electrical impulses in the proper sequence. The choice of one type of transmitter over another depends upon factors discussed below.

The single head transmitter distributor shown in Figure 5 is not used widely in tape relay operation although it does find application in certain off-line* functions and in branch offices, as subsequently explained.

The triple head, or three-gang transmitter distributor, Figure 6, is used for transmitting into wire or radio circuits either directly or through code converters. The three transmitter heads actually key the same channel in turn, operating one at a time. They are provided in order to hold to a minimum the separate operator functions and waiting periods necessary in the transmission of a series of messages. The head on the left is restricted in function to the sending of channel numbers and is normally fed with a continuous tape containing a series of numbers properly prefixed for the outgoing channel. The remaining two heads are for transmission of message tapes and are wired for automatic sequential or tandem operation. Thus, a second message

* Off-line designates those message processing functions which are performed on local apparatus dissociated from wire or radio channels, as distinguished from on-line functions which directly involve the handling of messages over wire or radio channels.

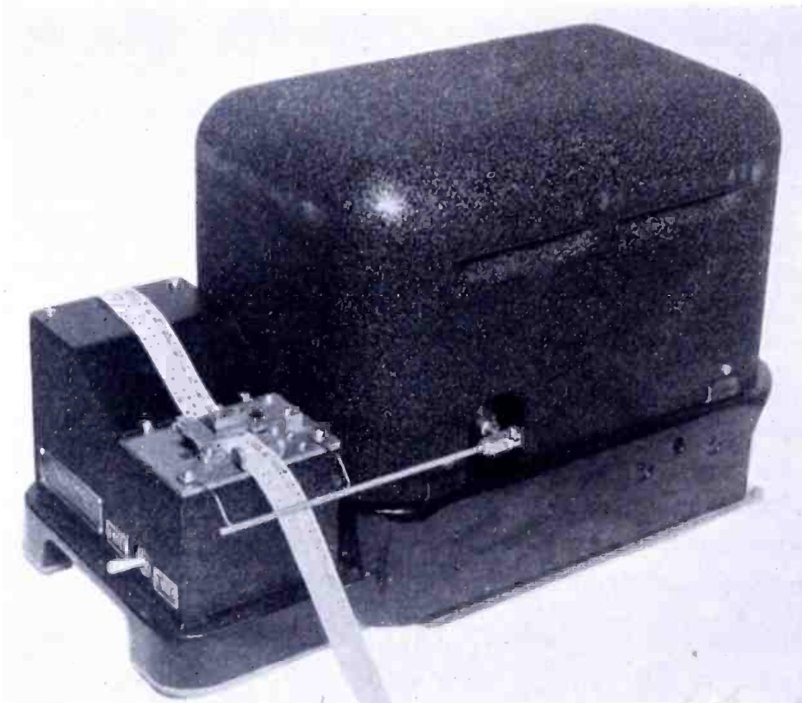


Fig. 5—Single head transmitter distributor. (Switch-arm in foreground automatically stops the transmitter if tape becomes tangled or taut.)

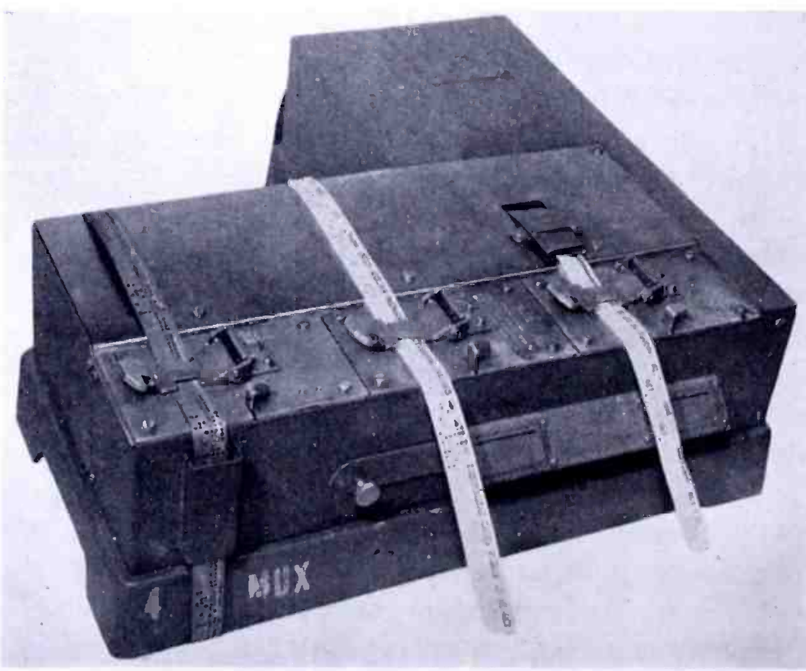


Fig. 6—Three-gang transmitter distributor. (Note heavy duty numbering tape which can be re-used many times without developing weaknesses or imperfections.)

may be placed in the transmitter distributor and made ready for transmission prior to the completion of the first. No action will take place until the first tape has cleared the tape-out pin, after which, without operator attention, the next channel number will be sent and transmission of the waiting message will proceed. The tandem type of tape feed takes place without any loss of time between messages, thus permitting full utilization of wire or radio facilities with a minimum of operator attendance.

The six-gang transmitter distributor, Figure 7, is used where it is possible and desirable to control several circuits from one operating position. It finds its major application on land-line channels. In order to effect the greatest practicable concentration of sending positions, the associated automatic numbering heads are mounted separately in identical six-gang assemblies.

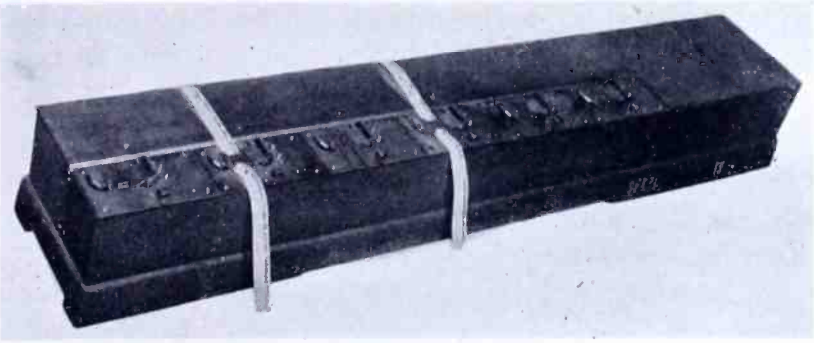


Fig. 7—Six-gang transmitter distributor.

Operating Assemblies—The Tape Relay System is built around the two basic devices described above, together with page printers for use at points of message termination and for certain central office functions such as monitoring. Basic elements are assembled in various combinations and mounted to suit each particular need. A typical example of functional mounting is found in the receiving reperforator console (see Figure 12) which accommodates eight receive-only typing reperforators in a compact group along with necessary number sheet holders, spring tape holders, and other supplementary components. The specific application of each device will be covered in detail in later sections of this paper.

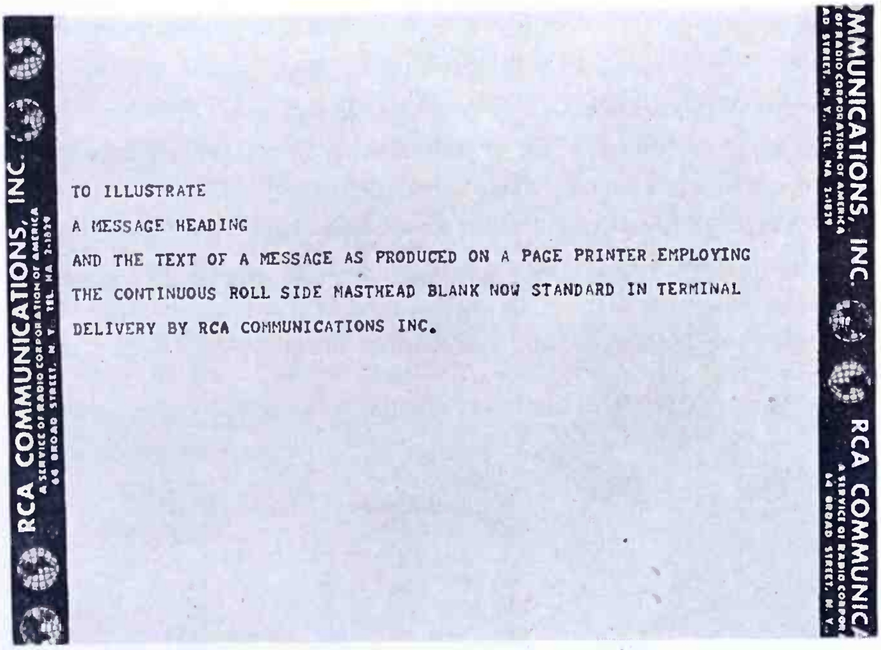
Processing Facilities for Originating and Terminating Traffic

In the Tape Relay System, messages are processed manually only at points of origin and are prepared in page form only at terminating points from which ultimate delivery is made. At all intermediate points messages are handled only in the form of typed reperforated tapes.

Originating Message Facilities in Branch Offices—For the processing of outbound traffic, each branch office is furnished with a send-receive typing reperforator for the manual processing of messages into tape form. Situated immediately adjacent to the reperforator is a single head transmitter distributor electrically connected with the central office. Tape flows in unbroken sequence from the reperforator directly into the transmitter so long as traffic is presented for transmission. During the interim period of changeover from Morse to printing telegraph tape relay operation, the larger branch offices are provided with dual lines for selective transmission, (a) to typing reperforators where messages are to be retransmitted over five unit channels, or (b) to page printers where manual retransmission is necessary, as in Morse or seven unit printer operation. The primary reason for use of the tape and transmitter distributor method of transmission rather than direct transmission from the keyboard of a send-receive page printer lies in the relative ease of correcting operator errors and resultant economy of circuit time and neatness of delivered messages. When the perforating operator strikes a wrong key he manually back spaces the tape, overstrikes the wrong perforation with the letters shift combination (all five perforations) and then resumes perforation. The letters shift combination is not printed in the page printer copy at the point of final delivery. Thus a minimum of circuit time is absorbed in the correction of perforation errors and the addressee receives a neat message copy.

Terminating Message Facilities in Branch Offices—Page printers electrically connected with the central office are used in branch offices for the reception of inbound messages. The center masthead type of receiving blank formerly used for final preparation of messages for delivery has been superseded to a large extent by continuous roll blanks with side mastheads, a sample of which is shown in Figure 8. This simplifies the problem of message spacing and eliminates the danger of having part of a message appear in the masthead area. At the same time it speeds up the entire process and permits non-continuous attendance of the terminal printer without danger of mutilations. Figure 9 shows a typical branch office equipment installation including the originating and terminating equipment.

Where it is necessary or desirable to prepare messages for delivery on center masthead blanks, it has been found advantageous to receive on a typing reperforator and prepare the final page copies on an off-line basis with a single head transmitter distributor and page printer mounted close together so that the operator can manipulate them as


RCA COMMUNICATIONS, INC.

TO ILLUSTRATE
A MESSAGE HEADING
AND THE TEXT OF A MESSAGE AS PRODUCED ON A PAGE PRINTER EMPLOYING
THE CONTINUOUS ROLL SIDE MASTHEAD BLANK NOW STANDARD IN TERMINAL
DELIVERY BY RCA COMMUNICATIONS INC.

RCA COMMUNICATIONS, INC.

Fig. 8—Side-masthead delivery blank.

required to produce suitable copies. If center masthead blanks are used for direct reception, an operator must be in continuous attendance,

Fig. 9—A typical branch office installation. (The equipment for perforating and transmitting outgoing messages is shown on the left, and that for receiving incoming delivery copies on the right.)

and sufficient time must be allowed between messages to permit blanks to be removed, inserted and adjusted.

Originating Message Facilities in Central Offices—The function of preparing messages for original transmission in central offices is identical with that in branch offices with the exception that, depending upon the physical arrangement and size of the office, tape may be carried physically rather than electrically to the sending position. In such cases the perforating operator is usually provided with a spring tape-holder into which each message tape is hung as completed. Clerks then collect

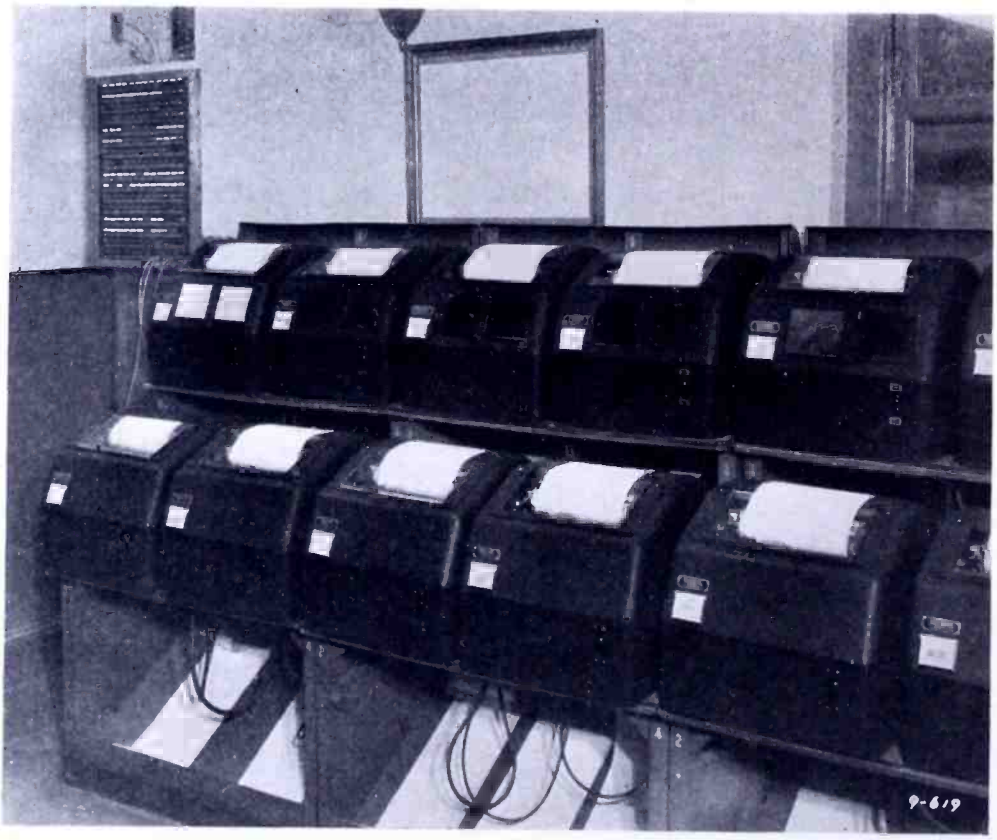


Fig. 10—Multiple printer mountings. (Four printers are mounted on each table; the jack-panel in the background permits ready rearrangement of line and printer connections.)

the tapes, route them and carry them to the proper circuits. Traffic accepted in page form from the domestic telegraph company, from tie-line facilities, from the telephone recording section or by physical transfer from branch offices is prepared in a perforating pool situated on the operating room floor in the same general area with the radio channel terminations.

Terminating Message Facilities in Central Offices—In the Tape Relay System, messages are not normally prepared for delivery in central office operating rooms; instead, separate delivery departments are located on street levels and operated on the same basis as branch offices. These delivery departments are equipped with receive-only page printers controlled from automatic sending tables in the operating room. In some cases several printers are required and in such situations multiple mountings of the kind shown in Figure 10 are used.

Local Distribution Facilities in Central Radio Offices

Outbound messages are received in, and inbound messages are delivered from Central Radio Offices in a number of different ways. Branch offices pick up and deliver the major share of the traffic. Plans have been developed for transferring hinterland traffic to and receiving it from the domestic telegraph company on the same basis as in the case of branch offices. Messages for large volume users are handled over private connecting printer circuits, and some messages are delivered by telephone over leased lines or regular commercial telephone facilities. The local distribution facilities for these different types of operation are described below.

Automatic Sending Tables—Pictured in Figure 11 is a bank of automatic sending tables over which most wire line transmissions are accomplished. Transmitter distributor heads in central offices are connected with printers in branch offices, in the telephone section, the service department and with the printers for the domestic telegraph company.

Each automatic sending table mounts two six-gang transmitter distributors which may be employed for simultaneous transmission over as many as twelve channels. Where traffic loads are sufficiently heavy, two adjacent heads in the six-gang transmitter distributors are connected to send on an alternate or tandem basis into the same land-line circuit. The two heads so used are usually located one above the other. Tandem operation permits a message to be made ready for transmission while another message is being sent, thus eliminating loss of channel time between messages and insuring efficient utilization of personnel.

Automatic channel numbering equipment used in connection with automatic sending tables is generally situated in the same area but not necessarily immediately adjacent to the message transmitter distributors.

Unsent traffic is stored at automatic sending tables in slotted racks, called washboards, in proper chronological order according to precedence classification. After passing through the transmitter distributor heads the tapes flow into a bin in the rear of the table from which they are subsequently removed and destroyed. A page printer monitor copy is made of every message sent to branch offices, and these copies are filed in numerical order according to the circuits over which they are received.

The primary functions of the sending operator are to scan tapes for correct routing and insert them into the proper transmitter dis-

Fig. 11—Automatic sending tables in operation.

tributor heads. The necessary functions of numbering outward messages, and preparing central office monitor copies are accomplished automatically without special action on the part of the operator. Line feed combinations included in the continuous number tape roll assure adequate spacing of messages at points of final printing.

Receiving Reperforator Consoles—Traffic from all points where messages are prepared for original transmission by the Company's employees and which are connected by wire with the Central Radio Office is received on receive-only typing reperforators compactly mounted in consoles of the kind illustrated in Figure 12. Each console accommodates eight receive-only typing reperforators and includes facilities for feeding out completed messages, for storing tapes, and for mounting the channel number sheets. A bulls-eye lamp blinks in response to incoming signals, thus providing protection against undetected equipment failures. This kind of mounting permits a single operator to

Fig. 12—Receiving reperforator consoles in operation.

attend a number of incoming channels, the actual number being dependent upon traffic volumes.

Private Line Printer Concentrator—In order to effect the fastest and most economical transfer of traffic between the Central Radio Office and offices of heavy volume users, private lines are provided to extend to such customers direct printing telegraph connections. Because it is impracticable to employ a separate machine for the central office termination of each private printer line, the lines are connected through a concentrator switchboard so arranged that any one of a number of sending or receiving positions may be quickly connected to the line upon receipt of a call from a customer, or upon presentation of traffic for a customer.

Facilities for Receiving from Private Printer Lines—A momentary opening of the line accomplished by pressing a special key on the printer at the customer's end locks in a signal lamp in the concentrator switchboard. The attending operator cords the line to an idle receiving printer. This action operates the motor-start relay in both the customer's and the central office printers which are otherwise at rest. The starting of his printer motor indicates to the customer that his line has been connected and that he may proceed with traffic. Normal central office reception from private printer lines is on page printers because few customers can or will prepare messages in the proper form for direct tape transfer to radio circuits. Facilities are provided, however, for cording in typing reperforators for press and other customers who are in a position so to prepare their traffic.

Facilities for Sending to Private Printer Lines—Messages intended for delivery over private lines are normally presented in typed reperforated tape form, just as they are received from distant stations. The equipment for sending to private printer lines consists of a single head transmitter distributor connected in series with a monitoring send-receive page printer. The operator receives the message tape with the customer's channel number and the routing marked with pencil on the leading end. He selects an idle sending position, has it corded to the customer's line and transmits the customer's channel number from the printer keyboard. He then starts the tape in the transmitter distributor and observes the page printer monitor copy for assurance of unmutilated transmission. Upon receipt of the customer's acknowledgment the concentrator switchboard cord is removed and the monitor copy bearing the customer's acknowledgment is retained for the central office record. A general view of the private printer line section presently installed in the Central Office in New York is shown in Figure 13. Active engineering study is in progress to mechanize the process of establishing connection in either direction, thus eliminating the manual switchboard operation and assuring faster, more accurate service.

Private Telephone Tie-Lines—Customers may dictate outbound messages over the telephone if they so desire, and in some situations, the Company makes telephonic deliveries. Where the volume of a customer's traffic warrants, he is given a private telephone tie-line which is terminated in a multiple turret switchboard in such a way that each line is accessible to a number of operators, one at each turret. Incoming calls are indicated by signal lamps in each of the several turrets, and outgoing calls to customers are accomplished by direct ringing.

Radio Circuit Operating Facilities in Central Radio Offices

Types of Radio Circuits—Circuits range in variety from those operated during a number of scheduled daily periods and handling only a small number of messages per day to circuits operated continuously and requiring a multiplicity of channels to handle several thousands of messages per day.

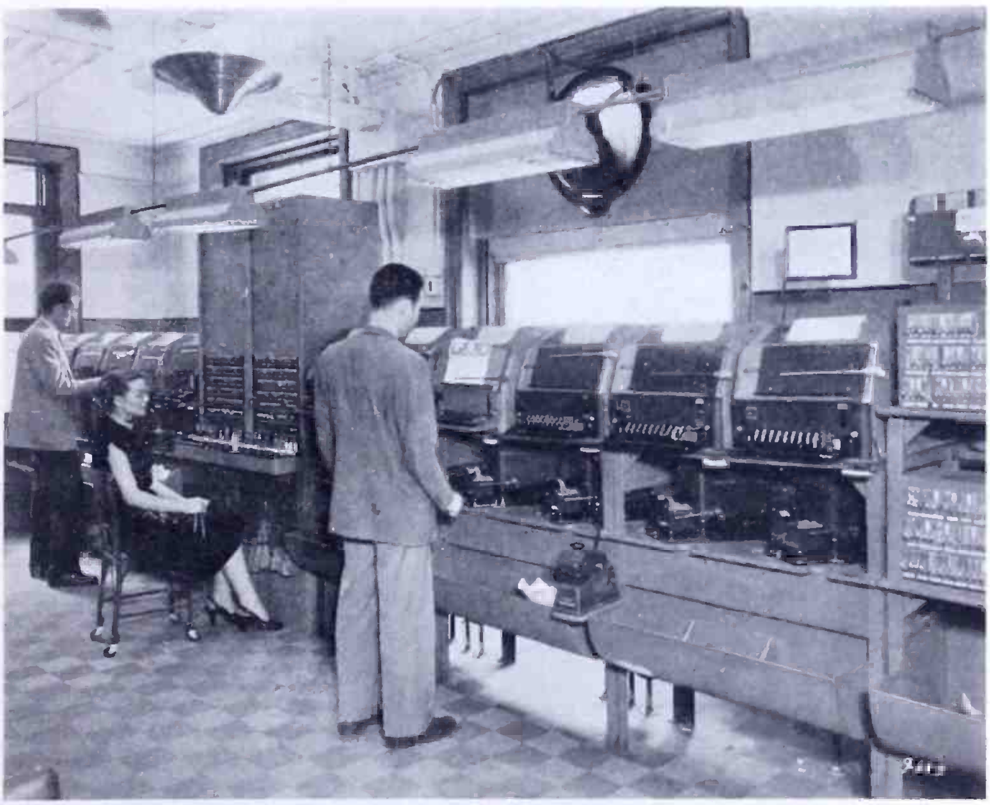


Fig. 13—Transmitting equipment and concentrator switchboard in the private printer line section, Central Radio Office, New York, N. Y.

A single radio transmitter may be used for single channel transmission in turn to several points worked on a schedule basis, for single channel transmission in turn to several points worked on a continuous basis, for single channel transmission solely to a single point, or for multi-channel transmission to a single point or to two or more points simultaneously.

For lightly-loaded circuits, single channel printer transmission provides adequate traffic handling capacity in suitable form. In other cases, where the traffic volume to be handled exceeds the capacity of a single channel, RCA Time Division Multiplex is used to provide a multiplicity of channels which may be used for operation with five unit code equipment or with equipment employing the RCA Seven Unit Printer Code with its valuable error indicating feature, or for combinations, in pairs of channels, of both kinds of equipment.

Five Unit Code Radio Printer Operation—The basic components of transmitting and receiving equipment required for the operation of a radio printer channel are essentially the same as those for land-line service. The principal difference lies in the manner of mounting and grouping the components for most effective circuit control and staff utilization. The same equipment, with minor additions or modifications, may be used to key a circuit shared by two or more overseas points, to key a single channel circuit used exclusively with one overseas point, or to key the individual channels of a circuit operated on a multiplex basis. In every case the requirements are for reception of intelligence in the form of typed reperforated tapes and for transmission from such tapes. Supplementary refinements such as automatic numbering, tape monitoring, page monitoring, etc., may or may not be included, depending upon individual requirements.

The Package Set—The equipment assembly most widely used for termination of the five unit radio printer channels is known as the package set. This assembly embodies in one console all necessary elements for transmitting into and receiving from one duplex radio channel operated by the tape relay method.

The package set, illustrated in Figure 14, comprises the following:

1. a three-gang transmitter distributor for automatic numbering and tandem type message transmission;
2. a receive-only typing reperforator located immediately above the transmitter for reception of incoming messages;
3. a receive-only typing reperforator and motor-driven tape reel mounted in the top compartment for continuous monitoring of the transmitted signal to facilitate possible re-running of traffic;
4. a relay and wiring unit mounted in the closed compartment at the bottom of the set accommodating all necessary relays, control switches, wiring terminals, etc.;
5. a signal indicator, visual and aural protective devices, and tape feed-out control buttons mounted in the panel at the top of the set; and
6. Compartments for sent tapes and waste paper, and facilities for storage of unsent tapes, received tapes, numbering tapes and number sheets.

The package set with its automatic channel numbering feature is generally employed on heavily-loaded channels. The proximity of the sending and receiving equipment under the control of a single operator facilitates quick and effective circuit management. Send-receive typing reperforators for the preparation of supervisory and operating instructions, are provided and mounted separately in the ratio of approximately one for each two package sets.

Since both incoming and outgoing message tapes are handled through package sets and so are apt to be confused with each other,

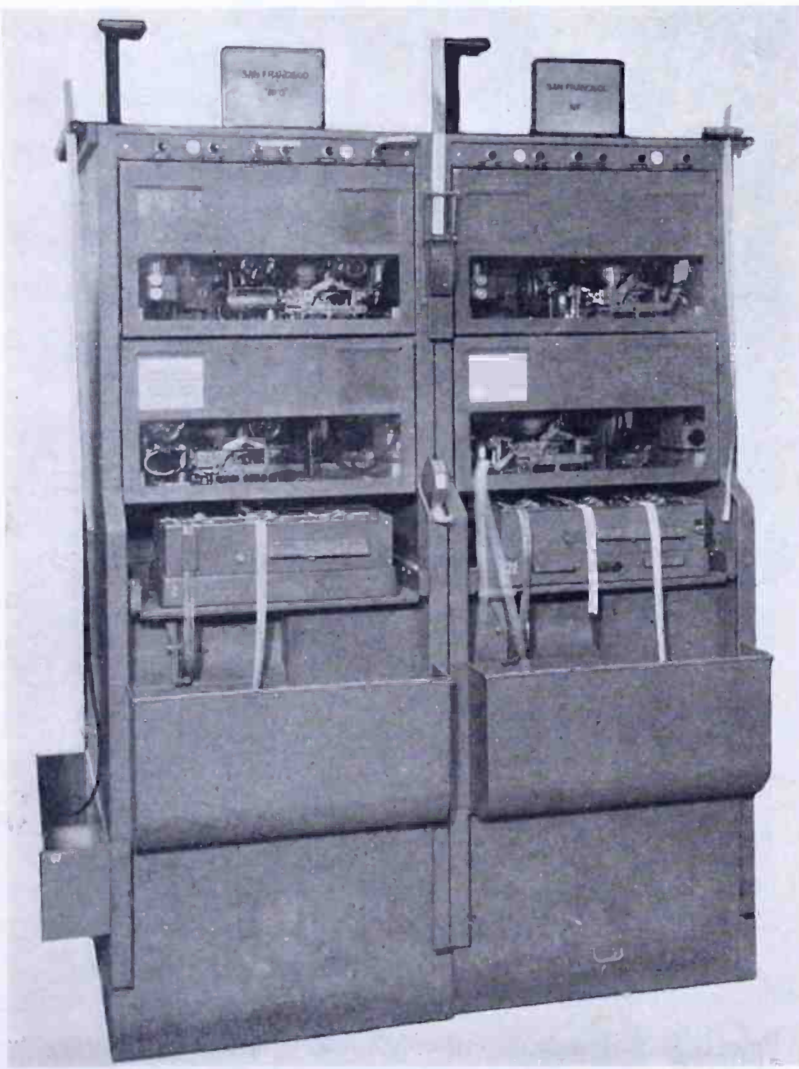


Fig. 14—A pair of package sets on two of the New York—San Francisco channels. (Each unit occupies a floor area two feet square.)

green tape is used for incoming messages to distinguish them from outgoing messages which are always white. Green tapes are never transmitted from package sets. A special procedure for handling transit messages is described elsewhere in this paper.

Radio Transmitting Consoles—To accommodate lightly-loaded radio printer circuits often operated on a shared basis with two or more points, a special table has been developed to mount a single head transmitter distributor, a monitoring typing reperforator, a selective system for controlling the motor start relays of two or more remote receiving stations and the necessary facilities for storage and disposal of sent tapes. Figure 15 shows one form of table adapted for this service.

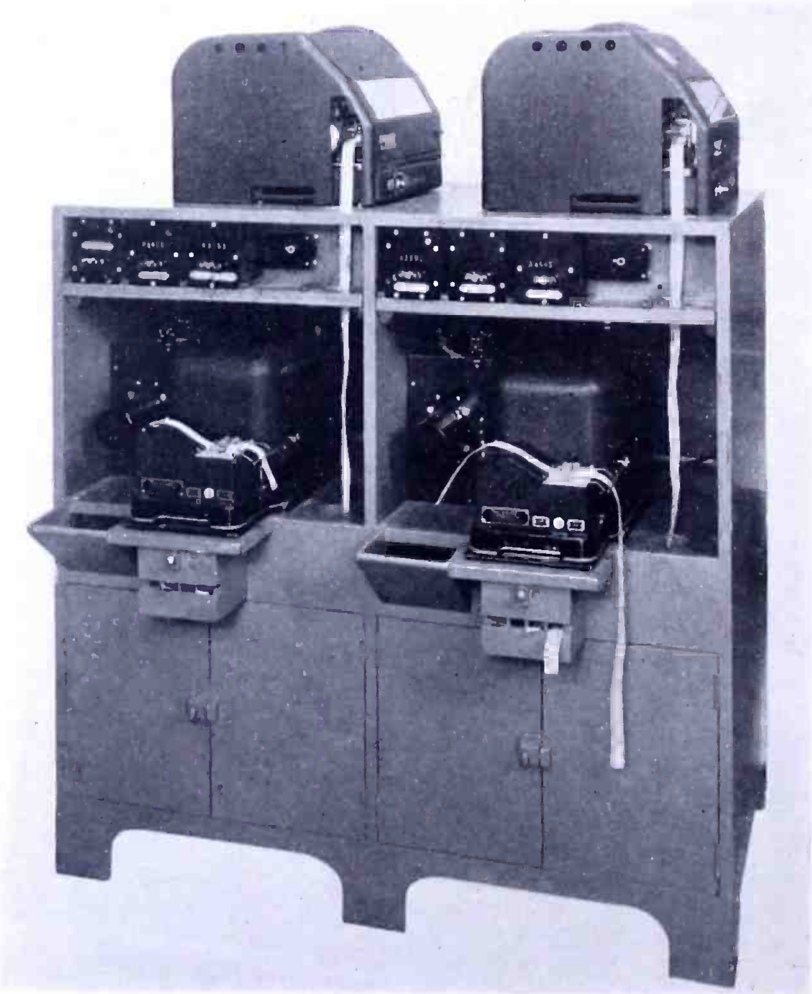


Fig. 15—Typical radio transmitting console. (Duplicate mounting controls two radio transmitters; each side is equipped with two motor-start selectors and one common motor-stop selector.)

Radio Receiving Reperforator Consoles—The receiving sides of channels operated from tables of the type described above are normally terminated in receive-only typing reperforators mounted in receiving reperforator consoles of the kind described above under **Local Distribution Facilities in Central Radio Offices.**

Accessory Facilities

The Tape Patching Set—Tapes which are physically damaged after reception, or those which contain obvious errors can be repaired locally by means of a tape patching set of the kind shown by Figure 16. The damaged tape is run through the transmitter distributor up to

Fig. 16—Tape patching set.

the point where the fault appears and then the transmitter distributor is stopped while the faulty section is reperforated manually on the keyboard of the send-receive typing reperforator. When the faulty section has been corrected, the tape is manually moved ahead in the transmitter distributor as far as necessary, and the remainder of the tape is reproduced automatically, thus providing a good tape suitable for onward transmission.

The Pneumatic Tape Tube—These tubes were especially developed for the purpose of quickly and economically transferring chadless tapes from one section of the operating room to another. No containers are required for the tapes, which are simply formed into loops with the raised chads on the inside and inserted into the tube gate as shown in Figure 17. Air is kept moving through the tubes by a suction pump located in the bottom compartment of the receiving terminal shown by Figure 18. Message tapes are drawn through the tube and deposited in the plastic covered receiving compartment as illustrated.

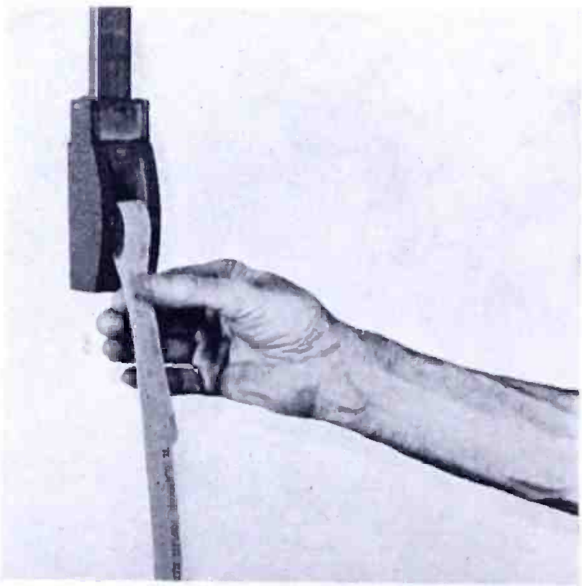


Fig. 17—Pneumatic tape tube gate.

The Spring Tape Holder—For the convenient storage of live message tapes, spring tape holders of the kind shown in Figure 19 are used in the Tape Relay System. As many spring tape holders as necessary are installed at traffic storage points, and message tapes are stored in chronological order from left to right and removed for transmission in the same order. As the storage and transmission of tapes proceeds to the point where convolutions at the right end of the spring are filled, the convolutions at the left end will normally have been emptied through transmission of tapes, and storage can again begin at this point.

Spring tape holders are inexpensive and experience indicates that they should be provided on a generous basis at all traffic storage points.

Tape storage springs should, where circumstances permit, be located above the top level of surrounding apparatus so that accumulations of unsent message tapes will stand out conspicuously.

It usually is advantageous to locate a trough of generous proportions several feet beneath the tape storage spring assemblies in order to prevent the dangling ends of long tapes from lying on the floor.

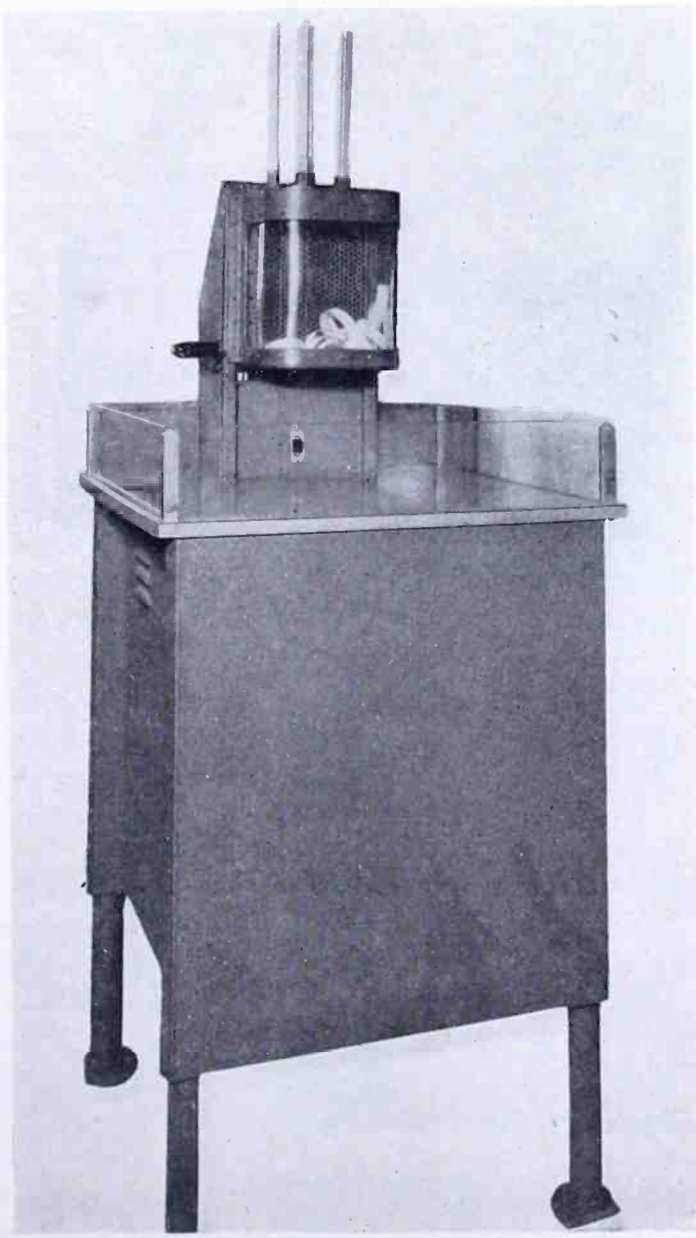


Fig. 18—Pneumatic tape tube receiving terminal. (A single terminal may accommodate up to six feeder tubes.)

Traffic Handling in the Tape Relay System

Information so far presented has been limited largely to descriptions of the components which make up the necessary equipment array for the tape relay system. The flow of traffic through the system and the control and protective measures applied to it are the points of primary concern and are dealt with in the following paragraphs.

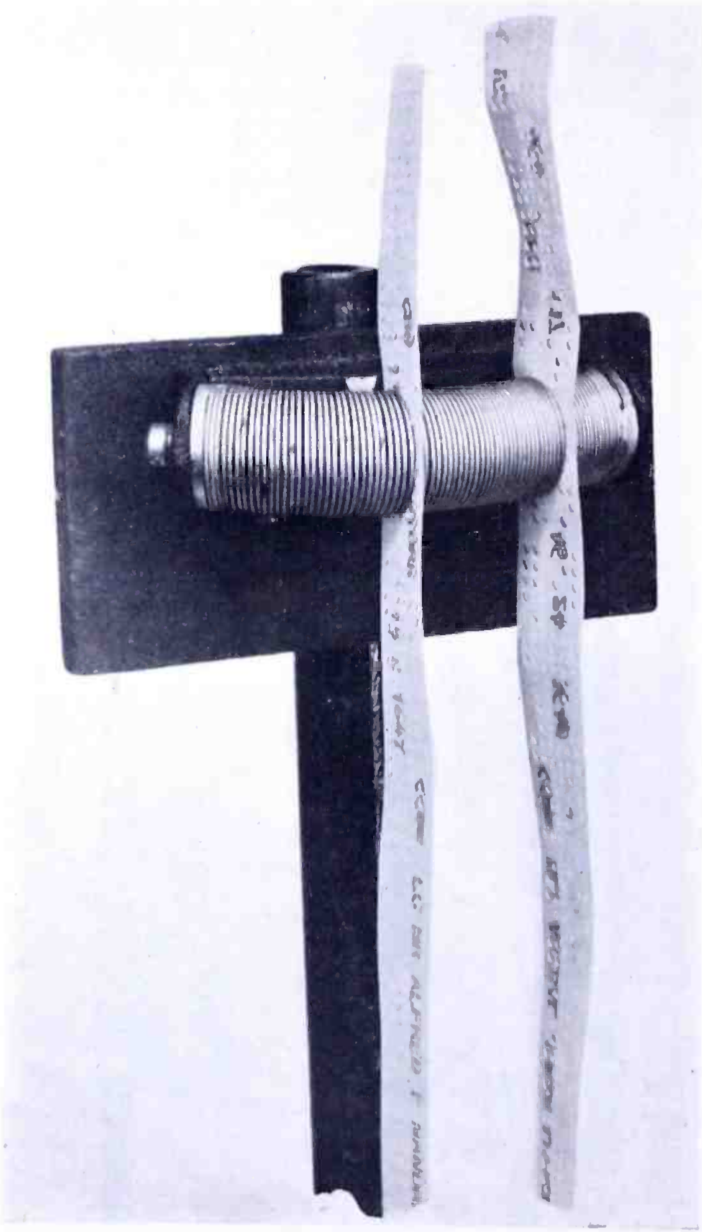


Fig. 19—Spring tape holder.

Original Processing of Outbound Traffic—Outbound traffic is prepared manually for initial transmission and introduced into the system at branch offices, or in central office perforating pools. All such traffic is assigned its circuit routing and original identification number prior to perforation and this information is perforated into the leading end of the tape. Traffic which originates in branch offices reaches the operating room through receiving reperforator consoles and is hand distributed to outgoing radio channels in accordance with the routing indicators which appear as the first characters on each tape. Traffic prepared for original transmission in the central office is collected from the perforating pool and distributed in the same manner as that received in perforated tape form from branch offices.

Message Numbering—A control feature of major importance in the successful operation of the Tape Relay System is the use of channel numbering. A basic identifying number, assigned without consideration of the point of destination, is perforated into the original tape by the branch or central office operator. The sequential numbers to any given point must be assigned by channel with a separate series for each channel, independent of message classification. Wherever practicable this is accomplished automatically. Channel numbering permits the immediate relay of message tapes received from widely scattered points without re-processing or delay. Channel numbering also reveals any losses of messages in transmission.

Monitoring—In order to maintain an unbroken file of messages on all channels, a page monitor printer is continuously connected to each transmitting channel of every character both in the local distribution system and in the radio tape relay system. Inasmuch as every message tape handled through the system must ultimately be transmitted electrically from the central office, universal monitoring of every transmission regardless of direction assures the presence of all messages in page form in the file section.

Transit Traffic—Messages which pass through the central office in transit from one radio channel to another constitute in a measure an exception to the monitoring rule in that separate copies of transit messages are required for the inward and the outward circuit files. Transit message tapes are, therefore, transmitted through a local transfer circuit from the automatic sending table to the receiving reperforator console in order to provide the necessary copies for the inward circuit files. A second purpose is served in the transformation of the message from inbound green tape to outbound white tape. The differentiation in color, as previously explained, is employed to prevent inadvertent mixing of tapes where inbound and outbound facilities are in close juxtaposition. The practice of never transmitting green tapes through radio transmitter distributors prevents misrouting.

Inbound Traffic—The Central Radio Office, insofar as received traffic is concerned, actually functions as a relay point with no terminating activity whatsoever. Thus each tape removed from the receiving reperforators on radio circuits must be electrically transmitted to its point of termination. The first step in the process is to deliver all such inbound tapes to a routing section for the interpretation of code addresses and indication, by pencil marking, of the proper route over which each message will reach its point of destination. Insofar as practicable, radio receiving positions are connected with the routing section by pneumatic tape tubes. Individual message tapes inserted in these tubes by receiving operators are delivered almost instantly to the routing clerks.

At each routing desk, a rotary card file system permits rapid translation or unpacking of code addresses, indicates the method and route of terminal delivery and reflects any special instructions for Sunday and holiday disposition. The routing is marked on the leading end of the tape which is then carried by continuously circulating clerks either to the automatic sending tables or to the private line printer section for onward transmission.

Messages frequently are in the hands of the delivering agency or, in the case of private line offices, in the addressee's hands in a matter of two or three minutes after receipt in the central office. Even under conditions of peak-load congestion, traffic flows through the system with negligible delay. Security of traffic is assured because losses in transmission are immediately apparent in the interruption of the channel number sequence and any loss or misdirection of a message tape in the central office is revealed when the monitor copies are assembled in accordance with the various radio channel number series.

Since it is impracticable to make a letter-by-letter examination of message tapes during transit through the central office it is possible for minor mutilations, wrong checks or collation differences to remain unnoticed until the message is received in page form by the delivering agency. Due to the very rapid relay process, however, the added time difference between transmission and query for correction is not appreciable. Meticulous check by the delivering agency tends to prevent errors from being carried through to ultimate addressees. In the case of tie-line transmissions, the sending operator carefully scans the message tape prior to transmission and observes the monitor copy during transmission to disclose any undetected errors. A comprehensive series of training manuals has been prepared for operators and traffic clerks which deal with this problem.

Radio Telegraph Operations Requiring Adaptation for Five Unit Tape Relay Service

Policy

It is the policy to convert operations on all circuits to the five unit code tape relay basis as expeditiously as possible, and for that purpose to utilize five unit code printing telegraph equipment directly to the fullest practicable extent. In some cases, however, extensive improvement of radio transmitting and receiving facilities is a necessary prerequisite to institution of printing telegraph operation of any kind. In other cases, correspondents have yet to equip their terminal offices with the appropriate printing telegraph and accessory equipment, so that the desired method of operation can become possible.

Pending complete changeover of operations on such circuits from the old methods to the new, it is the policy to make such adaptations as may be possible at its own terminals, since otherwise both incoming and outgoing messages must be reprocessed there letter-by-letter for forwarding. The extra handling time and cost involved in such reprocessing and the added opportunity for human errors are strong inducements for making adaptations where, as in central offices, the great preponderance of traffic can be collected from and delivered to branch offices and hinterland points, and delivered to tie-line customers most effectively by five unit tape relay.

Circuit operations of types requiring adaptation are described below.

Error Indicating Seven Unit Printer Code Operation—The valuable error indicating properties of the RCA Error Indicating Seven Unit Printer Code accrue from the fact that the use of seven signal elements per character permits the assignment of code combinations in such a manner that each character is represented by a different combination of three marking and four spacing signal elements as shown by Figure 20. The receiving instrument is so designed that if more or less than three marking signal elements are received during the time interval allotted to each complete character, the resulting false combination is recorded as an error.

The currently available seven unit code equipment is of the type which depends upon external means for holding the receiving instrument in synchronism and phase with the distant transmitting instrument. Because of such equipment limitations, utilization of the seven unit code has been restricted to circuits provided with RCA Time Division Multiplex equipment or with other single or multiple channel equipment which likewise includes appropriate synchronizing, phasing, and signal "regenerating" means.

Further, the current seven unit code equipment does not include the typing reperforators, automatic numbering transmitters, etc., required for effective tape relay operation. Consequently, inbound messages handled over seven unit code channels are currently received in the form of printed gummed tape which is pasted to message blanks. Correction of any errors appearing in the initial reception is effected by having the mutilated text retransmitted and pasting the piece of tape containing the repetition over the original tape. In order to retain a central-office file copy, an impression or Ditto duplication process copy is made before the message is released for forwarding.

Messages received as described above must be manually reprocessed letter-by-letter for subsequent retransmission. Similarly, any outbound messages reaching the Central Office in the form of five unit reperforated tapes are not retransmitted directly from that tape. Instead they must be reproduced in page form and then be manually reprocessed letter-by-letter for transmission from unprinted seven unit per-

7 UNIT PRINTER CODE

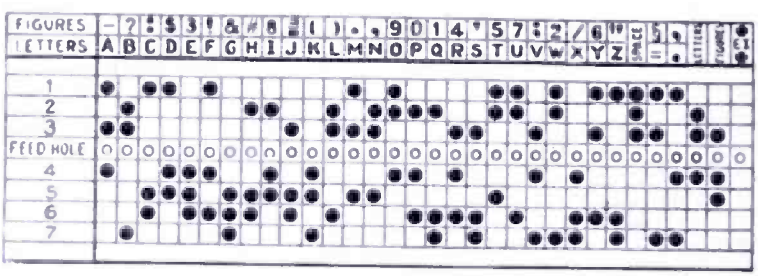


Fig. 20—RCA error indicating seven-unit printer code.

forated tape, except where five to seven unit interim code converters are used.

Means for coordinating use of the seven unit code on radio circuits with use of the five unit code for tape relaying, through the medium of code converters, are described elsewhere in this paper.

Higgitt Operation—The Higgitt or DCCC printing telegraph system is based upon the double current cable code, which is a polar version of the Morse code. Two channels are imposed on a single radio carrier with speed of operation adjustable between a nominal 40 and 100 words per minute per channel, depending upon the stability of the radio carrier. Reception is in the form of printed gummed tape and transmission is from unprinted Morse perforated tape so that, as in seven unit code operation, extra processing time and cost are involved in Higgitt operation. There is, however, the difference that non-typing reperforators are available for reproducing cable code signals in unprinted Morse perforated tape form suitable for retransmission over outgoing Higgitt or Morse channels.

Morse Operation—Morse operation, as used in most central radio offices, employs inked slip recorders for receiving. These reproduce Morse code combinations in an undulating line on a continuous narrow tape. This tape is pulled mechanically across a bridge in front of an operator who manually transcribes the code into its typewritten equivalent, producing simultaneously an original and a file copy. Skilled Morse operators can transcribe at fifty to sixty words per minute. A sample of such recorder slip is illustrated in Figure 21. Messages transcribed from Morse tape as above must be manually reprocessed letter-by-letter for subsequent retransmission.

Some experience has been acquired in the direct manual transcription of Morse recorder slip into five unit tape, using send-receive typing reperforators instead of typewriters, and while this method reduces the speed of the transcribing operator somewhat, it has an

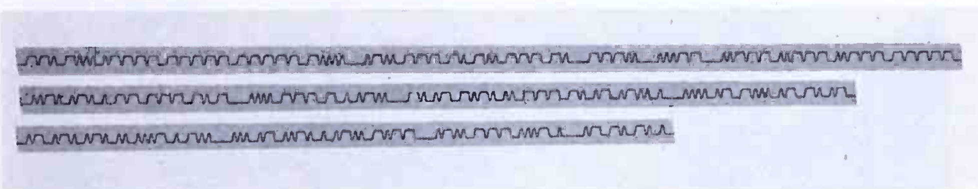


Fig. 21—Morse recorder slip. (The wide peaks represent dashes and the narrow peaks dots; the wide intervals represent word spaces and the narrow intervals letter spaces.)

overall advantage over typewriter transcription both from the point of view of ultimate speed of service, and that of economy.

Transmission is accomplished by passing Morse perforated tape through a Wheatstone transmitter. Speeds of keying the radio channel are adjustable over a wide range up to extreme speeds of several hundred words per minute, but most operators are limited to approximately sixty words per minute in the preparation of Morse perforated tapes. All messages to be transmitted in Morse must be manually perforated into Morse tape, letter-by-letter, using keyboard perforators except where five unit to Morse code converters are used.

Code Converters

In order to obtain the benefit of five unit tape relay operation to the fullest possible extent, despite continued use on various radio circuits of telegraph codes which do not permit direct reproduction of received messages in the form of five unit printed reperforated tape or direct transmission from such tape, a series of code converters has been developed for translating between the five unit code and the others. The five unit/Morse and Morse/five unit and five unit/seven unit code converters described in the following paragraphs are considered to be interim devices to be used only during the period of transition to five unit and seven unit code operation. The converters for translating both ways between five and seven unit code signals are designed with the purpose of continuing them in use, as the RCA Error Indicating Seven Unit Printer Code has advantages which the Company intends to retain and develop.

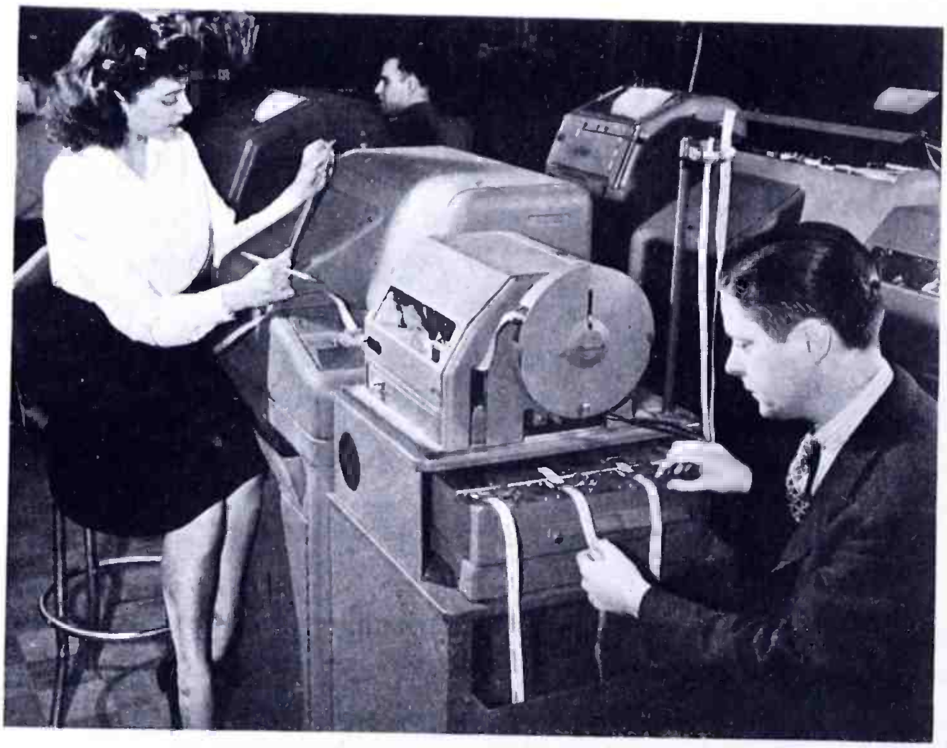


Fig. 22—A five unit to Morse code converter in operation. (Since more than one sequential number series may be processed through a single converter, expendable chadless number tapes are introduced in lengths required depending on the amount of traffic of a common series on hand.)

The Five Unit to Morse Code Converter—This converter comprises apparatus for automatically numbering and making page printer copies of messages that pass through it. Figure 22 shows such a converter in operation. The three-gang transmitter and associated receive-only typing reperforator mounted at the right hand end of the set are solely for the purpose of permitting automatic numbering. Except for this requirement, both of these units could be omitted, in which case five unit tapes would be inserted directly into the converter unit, from which corresponding perforated Morse tapes would emerge. Automatic numbering is accomplished in the same way as in package set operation. The monitor page printer shown at the left hand end of the set prepares a copy of each message translated. Morse tapes are separated into suitable lengths and physically transported to the proper circuits.

The five unit to Morse code converters described above require use of special procedures by operators perforating traffic which is to be translated. Most punctuation marks must be spelled out because the converter has no means of translating them into Morse signals, and

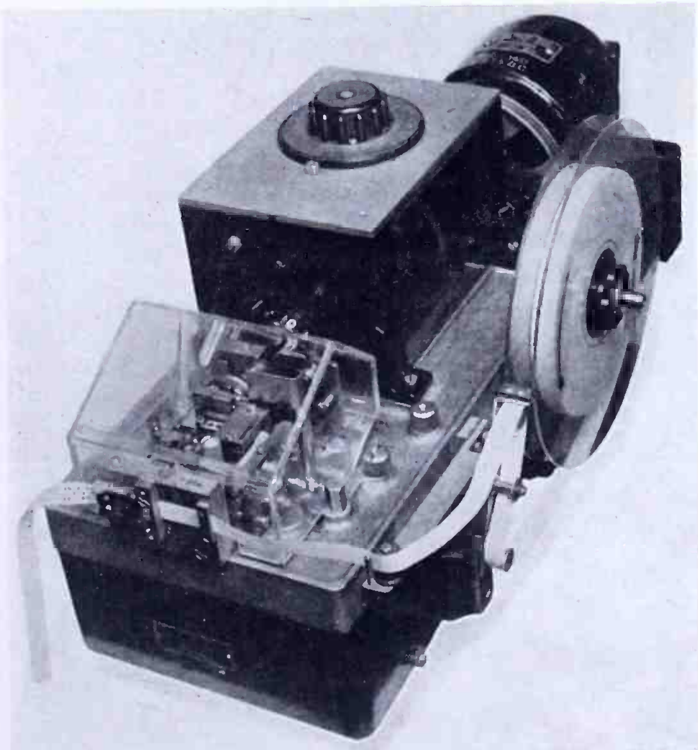


Fig. 23—A Morse reperforator with dust cover removed. (These devices produce unprinted, fully perforated tape corresponding to received Morse signals.)

the originating operator must send an upper case carriage return for the Morse "AR" signal, and a combination of one line feed and one carriage return for the Morse "BT" signal. These special procedures tend to slow down the operator to a considerable extent.

Morse to Five Unit Code Converter—This converter requires that Morse signals be received on reperforators, the tape from which can be inserted into converters. One type of Morse reperforator suitable for this purpose is shown by Figure 23.

A Morse to five unit converter of the type that has been used successfully is shown by Figure 24. Morse tapes from the reperforator are inserted in the Wheatstone transmitter and the messages they contain emerge in the form of five unit tapes from the typing reperforator. Since Morse tapes do not include the carriage return and line feed functions necessary for the proper operation of the page printer, these are inserted automatically by the converter.

Morse to five unit code converters are subject to no such limitations

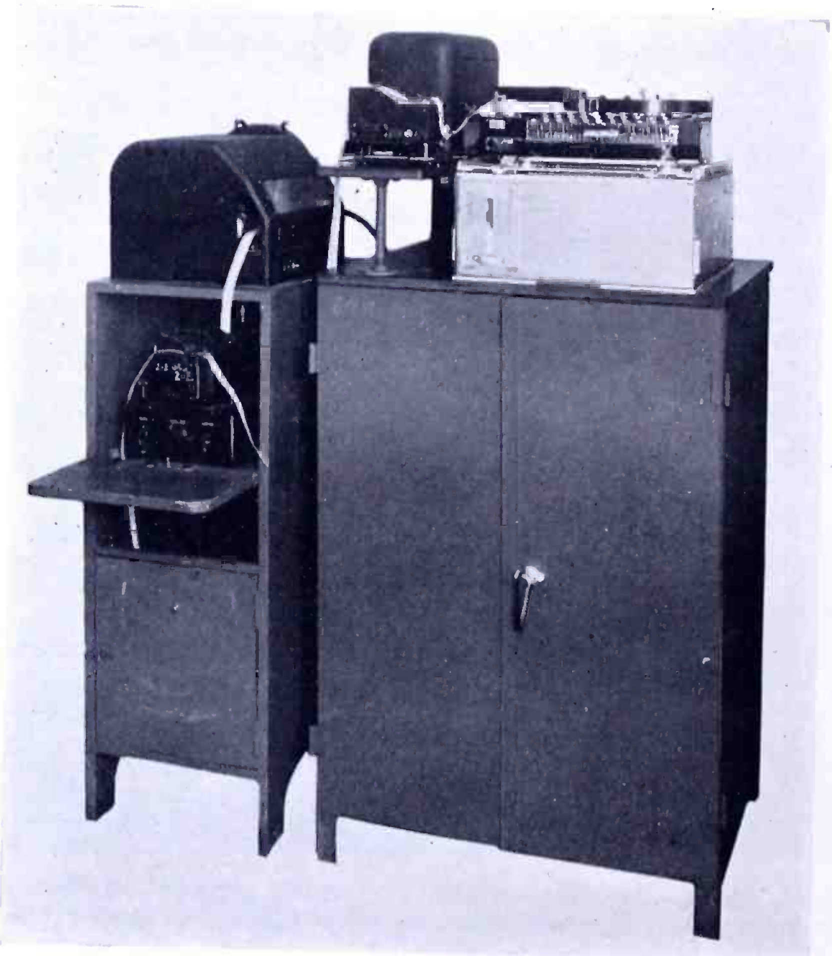


Fig. 24—A Morse to five unit code converter. (The translation and interim reperforation are accomplished in the element on the right which may be installed remotely from the operating elements on the left; the lower operating unit is the Wheatstone (Morse) tape transmitter and the upper unit is the five unit typing reperforator.)

as five unit to Morse converters because every character capable of being transmitted by Morse is translated by these machines into five unit code characters. The speed of operation of the Morse to five unit code converters is the principal drawback to their efficient utilization. They are presently capable of operating at only thirty-five words per minute and this generally necessitates the use of more than one converter on a single circuit.

Five to Seven Unit Interim Code Converter—In order to expedite the handling of messages outbound over seven unit code printer channels, an electro-mechanical converter of the type shown in Figure 25 has been developed. This device accepts five unit tape as received from

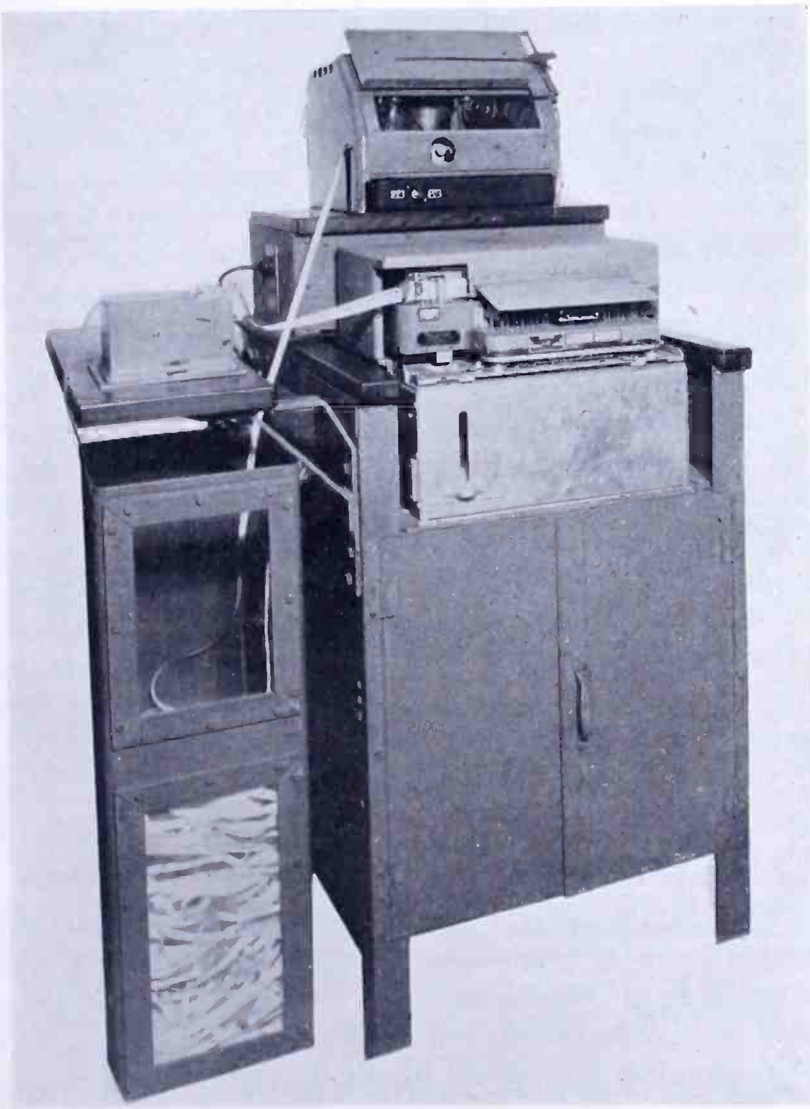


Fig. 25—A five to seven unit interim code converter with seven unit tape transmitter and tight tape switch permitting transmission into the circuit almost simultaneously with the conversion process.

land-line circuits and automatically produces a corresponding seven unit code tape.

An automatic numbering three-gang transmitter distributor is used for the transmission of messages from the five unit tapes into the converter. This transmitter distributor is located adjacent to the receiving reperforator consoles in which the receiving sides of land-line circuits are terminated. Operation of the three-gang transmitter distributor is the same as described in connection with the corresponding operation of package sets.

The converter is so designed that solenoid magnets operating in response to five unit signals control the key levers of a seven unit code perforator, the tape output from which is fed directly into a seven unit code transmitter.

Several converters of this type are in daily traffic service and perform with reasonable satisfaction, bringing to outbound seven unit code operation some of the advantages of tape relay and automatic numbering. Some disadvantage results from the necessity for operator attendance both at the five unit code transmitting position and at the converter, although one operator at each of these positions is able to control more than one circuit.

Five/Seven and Seven/Five Unit Code Signal Converters—With the purpose of retaining the error indicating feature of the seven unit code in its tape relay plan of operation, the Company has developed code converters capable of translating directly from five to seven and seven to five unit codes without interim reperforation. Thus, while retaining the flexibility of the fully developed five unit code tape relay equipment in terminal operations, error indication is achieved by transmitting a seven unit signal over the radio path. Because the conversion is on a signal basis, no additional operations are involved. These converters are in an advanced stage of development.

Extent of RCA Tape Relay Network

Major Tape Relay Stations

In addition to its Central Radio Offices in New York and San Francisco, the Company operates major tape relay stations in Tangier, Honolulu and Manila. Figure 26 shows the areas expected ultimately to be served through these stations either on a regular basis, or as an alternate arrangement when direct operation is interrupted or seriously impeded.

The relay station at Tangier has demonstrated its effectiveness by increasing quite materially the hours of operation with Bombay, and with Stockholm, Moscow and other European points. Printing telegraph operation over routes which pass close to the North Auroral Zone is not practicable much of the time, and for all radio-telegraph traffic centers the direct routes to which traverse this area, alternate routes via Tangier are advantageous. The facilities at Tangier are being steadily increased.

Honolulu and Manila are being developed to function in the same manner as Tangier, and their use will improve the service to Australasia and the Far East.

Special Services Made Possible by the Tape Relay System

Radio Printer Conference Service

This service permits direct two way printer communication between subscribers at widely separated points. Transmission normally is accomplished by a printer operator from dictation or from written

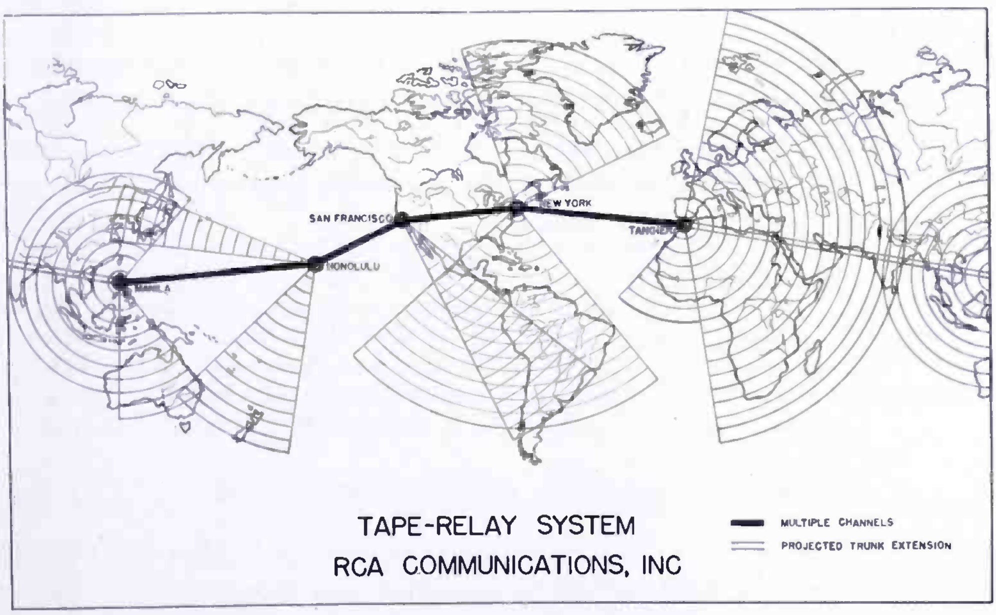


Fig. 26—Map showing the main trunk channels in the Tape Relay System and the areas of projected coverage for each of the major Relay Stations. (In addition to this network, direct channels are maintained between the United States and the principal cities of the world.)

material, while reception is by means of page printers, the copy from which is immediately accessible to the conferees.

Radio printer conference service was widely used by the Armed Forces during World War II with great success. This company, in cooperation with its correspondents, presently furnishes the service between the United States and Russia for the United States and Russian Governments, and between Lake Success, New York and Geneva, Switzerland for the United Nations. The service will be extended to other points as requirements develop and availability of facilities permits.

Volume Press Service

This is a fast and inexpensive service offered to newspapers and press associations. The subscribers must process the traffic into five unit tape form, and the intelligence is then passed through the Tape Relay System to the addressee with no manual letter-by-letter reprocessing at any point.

The rate for volume press service to any country is fixed on the basis of the normal press rate for some agreed-upon initial number of words in one direction in any month and a substantially lower rate for words in excess of this number. In every case a rate acceptable to the foreign correspondents is negotiated.

Conclusion

The Tape Relay System represents a new system of radio telegraph communications which provides the best service attainable with known methods at low unit cost. Service to points to which tape relay operation has been extended has proved to be greatly superior in all respects to that furnished by older methods.

COLORIMETRY IN TELEVISION*

By

WILLIAM H. CHERRY

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—*The colorimetrically exact reproduction of color in simultaneous television is now possible, through the congruence of the camera spectral sensitivities to definite characteristics specified by colorimetry and through the combination of the camera signals in both positive and negative amounts by suitable circuits and amplifiers. The negative sensitivities of the photo-pickups formerly required for certain spectral wavelength intervals are obviated by these signal mixing circuits, and simultaneous color television, both as to color range and accuracy of reproduction, is capable of the finest color reproduction available anywhere. The mixing circuits perform a function resembling that known in color photography as masking, but whereas the latter is always approximate and often a hit-or-miss procedure, the television system can approach perfection without undue complication.*

The basic concepts and relations of trichromatic colorimetry are here developed. Many of these relationships are of immediate importance in color reproduction and are stated explicitly, with the aid of a concise notation. In addition, certain rather philosophic aspects of television as a means for the communication of sense perception are discussed and a plea is made for the extension and compilation for television purposes of knowledge about the properties of the eye.

THE GENERAL PROBLEM OF COMMUNICATIONS

THE ADVENT of color television has been hailed by many as introducing a new 'dimension' in the field of communications. While the broadest possible interpretation is undoubtedly to be given to the word 'dimension', its easy and natural use in this connection suggests an underlying concept which is appreciated but vaguely. The literal-minded communications engineer who is unfamiliar with the methods of colorimetry† might well take this descriptive word as both accurate and unachievable. He is all too painfully aware that the medium available for communications is, and has always been, irrevocably two dimensional. The communication of all information must be accomplished in the relationship to time of some potential appearing at an input terminal or on an antenna lead. Of course, there are in principle several such relationships available, corresponding to the number of electrically insulated

* Decimal Classification: R583 × R800 (535).

† See REFERENCES on page 459.

channels used, thus the number of wires or, in wireless, the number of separate microwave or light beams. However, the limitations to finite bandwidth which appear in every system take away much more from the ideal communications medium than can ever be gained by the use of many insulated channels. Indeed, the correspondence of the communications medium to the information to be transmitted is much more like that of a limited region of the two dimensional space to the manifold of the information. How then, one may inquire, can the description of a real scene, even in black and white, where the light intensity, three spatial dimensions, and motion or time variation require a five-dimensional expression, be communicated through a less than two dimensional medium, still again such a description in color, which adds another 'dimension'?

The object of a television communications system is to provide to the receiving person the same sense perceptions he would experience were he present and observing the scene transmitted. This is a very important difference from the transmission and duplication of the physical specification of the scene, which is impossible. Not only are devices for reproducing some physical effects necessary to the purpose quite unknown, but, as has been indicated by the reference to dimensions, the amount of information to be conveyed is infinitely larger than the medium can transmit. Even so, the transmission of the physical data, at least to a crude approximation, is often thought of as the object of radio and television. As a result of this misapprehension, in the many inevitable engineering compromises choices are sometimes made which are very prejudicial to the true purpose of the system. An example of both the right and the wrong attitudes will appear incidentally in the subsequent discussion of color reproduction in television. However, first the answer must be suggested to the problem of dimensionality just now posed, and, as is well known, this answer is contained implicitly in the objective stated: the communication of sense perception, rather than of physical specification. Lest there appear a philosophical inconsistency at this point from the fact that physical specification is itself the formulation of sense perceptions consequent to the performance of operational experiments, it should be remembered that the number, type, and complexity of experiments carried out in one moment by the average spectator are far inferior to those required for that specification. Not only are all kinds of laboratory type experiments implied by the latter, but indeed an infinite number of experiments are in principle required for a complete specification.

Animal senses have but finite resolving power, and are sufficiently rough, crude, or insensitive that indefinitely large numbers of different sets of physical stimuli produce the same, or indistinguishably different, sets of sensations. Generally speaking, the sets of stimuli which are thus indistinguishable differ continuously in physical specification within an interval more or less small percentagewise[1]. Sets in a different interval must then produce a different set of sensations, and the inference is likely to be made that continuously variable causes, as would appear in the transition between adjacent intervals, produce discretely different sensations. This conclusion is inconsistent with the supposed property of finite resolving power of the senses, and the paradox has arisen through the tacit assumption of transitiveness of perceptual likeness[2], which is incorrect. Thus if a picture of 1000 lines cannot be distinguished from one of 10,000 lines, one of 900 from one of 1000, and so on, it does *not* follow that one of 200 cannot be distinguished from one of 10,000 lines. One might therefore argue that the same sense perceptions and indistinguishably different sense perceptions are not equivalent, but in a pragmatic sense such a debate would be profitless and quite immaterial, for the proper basis of comparison of a reproduction is always the original, and not some intermediary reproduction.

If, therefore, the television system presents a set of physical stimuli which, as perceived by the spectator at the receiver, is indistinguishably different from that as perceived by the same spectator when present on the scene transmitted, the system may be regarded as perfectly performing its function. It is thus necessary for the system to present any one set of physical stimuli corresponding to the finite interval of perceptual irresolvability in which the original lies, and since the total gamut of physical specification is itself finite, the information to be transmitted is reduced in quantity from the manifold infinite of the physical specification to correspondence to a finite number of points and is accordingly able to conform to the transmission medium.

The present bounds of technological development and economy have made it necessary to reduce the amount of information transmitted still further by a substantial factor beyond that following from the basic communications principle which has just been discussed. It is interesting to note the great variety of sense perceptions which have been considered of secondary importance and have been lopped off and forgotten altogether. Thus, only the spectator on the scene will

[1] In a certain sense the perception of color departs from this rule in that if properly chosen, the interval need not be particularly small, but this is of no consequence here.

[2] That is: if A is like B and B is like C, then A is like C.

perceive mechanical motion and vibration, atmospheric effects of wind, temperature, humidity, and pressure on the skin and ears, and the effects of odors, gases, dust, and so forth. As regards the two major senses with which communications systems are ordinarily concerned, sight and hearing, a great deal of information is also deleted: the absolute levels of the sound volume and of the scene brightness, the absolute scale of size, while the viewing angle is reduced to that of a small screen, and practically all perception of perspective. In the last category are included not only binaural and binocular perspective, but the sound and sight perspectives obtained through head motion and the visual perceptions obtained through the natural process of focus accommodation of the eye and its limited depth of focus. The reader may be able to add one or two items to this list of communications omissions, but in any case it is clear that present systems attempt merely to transmit a single sound signal and the visual signal associated with a simple flat image of the scene.

Once again, within the comparatively much smaller demands of a flat image, the overall engineering problem is to introduce a communications mechanism which, from this limited set of physical stimuli at the transmitter, reproduces one at the receiver yielding substantially the same set of perceptions. With the correlationships of physical stimuli which give rise to the same sensation playing so vital a part in the television problem, and governing all choices between technically different modes of transmission and presentation, is it not surprising that engineering knowledge of the visual properties of the eye is so incomplete? Indeed one might think that these properties would already have been so exhaustively explored, tabulated, and cross-indexed, that a handbook devoted to that subject, or at very least, a chapter in some standard communications handbook, would already be available and in constant use. It is not that there have been no comprehensive and reliable investigations of the phenomena of vision — quite the contrary, for a large amount of literature is available on the subject. However, the objectives in most of these investigations were sufficiently diverse and removed from the field of communications that the data are not suitable for direct application in that field. Certainly there has never been an adequate compilation and codification of that data for television purposes.

The physical specification of the flat image of the scene appearing before the television transmitter, apart from properties such as polarization, phase of the light waves, and so on, important in physical optics but, so far as is known, not directly entering into sensation phenomena, is a comparatively simple problem for mathematical description, and accordingly should be amenable to systematic investigation. One phase of such correlation is fairly well established in the science of colorimetry. But even here, as in general, two complex difficulties arise which may often be of secondary importance but are not always negligible. One is that what the eye sees in one part of the field of view depends to some extent on what is present in other parts, and the other difficulty is that any given eye changes its properties from time to time, under different non-visual stimuli, and after different previous experience, while different people's eyes, and even the eyes of the same individual, differ noticeably in their visual properties.

The first of these difficulties may be set aside to a certain extent by stating the visual data pertaining to one part of a field, under two headings, depending on whether the center of vision of the eye is directed to that part of the field or elsewhere. In most visual properties, flicker perception being a notable exception, it is found that the visual data are the more detailed and the more critical, and that the eye displays the greater resolving power, when the center of vision or center of interest falls upon that part of the field whose visual effects are being derived, than when it falls elsewhere. In addition, it is found within broad tolerances that the details of the other parts of the field have substantially no effect on the results and the rest of the field usually may be characterized in the large as background. In many visual properties even the background as a whole plays a very secondary role. Evidently then, these data pertaining to foveal vision, so-called because the image of the field of view at the center of interest falls on that part of the retina called the fovea, contain for the most part the information which is directly pertinent in television problems.

The second of these difficulties can be fairly well accounted for by establishing a statistical norm for the visual data collected from many different subjects, whereby it is found that the properties of the eye of the so-called standard observer thus obtained are satisfactorily representative of much of the data, and for engineering purposes the remainder is well taken care of when expressed in terms of probable deviations. This is the general procedure adopted in colorimetry, although data pertaining to deviations of individuals from the standard observer are rather sparse.

To illustrate the manner in which, it is hoped, visual characteristics may be correlated to the physical stimuli presented by a flat image, as, for instance, either that appearing before the transmitter or that at the receiver, consideration is given to the description of such images in terms of the light energy flux or intensity emitted, expressed as functions of wavelength, position, and time. The dependence of such intensity on direction of emission is of comparatively much less importance. While the physiological perceptions themselves are the consequence of the nature of these intensity distributions, the television engineer is primarily concerned with the visual differentiation between the reproduction and the original, and in this regard the various properties of the eye are functionals of these intensity functions. For purposes of measurement the simplest and most objective criterion applicable is that of distinguishability, but of itself this is insufficient because in the unhappy commonplaces of imperfection the measurements must be not merely of kind, but also of degree. Fortunately, in many cases it is quite easy to extend the criterion of distinguishability. For example, the degree of departure of the reproduction from the original image coming about through the line raster structure of the latter, and defects in the system, can be measured by the increased viewing distance necessary before the difference disappears. The magnitude of flicker in a reproduction can be measured by the reduced overall intensity amplitude at which the flicker disappears. Differences in color are capable of this same general examination, that is, by the number of intervening steps of barely detectable differences. It is important, however, to avoid assuming that there is a direct correspondence between measurements of this type and aesthetic values or secondary physiological effects such as annoyance and fatigue, because this is not always the case.

With further reference to the light intensity function, the usual categories of visual properties are accounted for accordingly as there are no time variations and the spatial variations are independent of wavelength (which covers monochrome visual acuity and resolving power); no time variations and no spatial variations except for perhaps one or two such to show matching (which includes most of colorimetry); no time variations but with spatial variations (to include color resolution and acuity); and with time variations but no space variations (for flicker, sequential color mixing, and after-image effects).

From the point of view of the television engineer the expansion of such light intensity functions in Fourier series in the space and time coordinates has direct meaning in terms of the electric signals which are transmitted, and it would be singularly convenient if the visual data, as, for instance, resolving power, could be expressed in direct relationship to the functional properties of such expansions. This is too much to expect, however, for some of the expansion terms would have negative values and thus be of doubtful physical significance when standing alone, while the usual sense responses are more often of a logarithmic nature. In other words, the visual properties are likely to have only very complex connections with the terms of a straightforward expansion.

As will be discussed in greater detail in the following section on colorimetry, the tristimulus system of color specification leads to the expression of most of the visual properties associated with the spectral distribution of a light intensity function by a set of three parameters, provided that spatial and temporal variations are sufficiently suppressed. These three parameters may be obtained by the wavelength integrations of the products of the spectral distribution with three functions obtained empirically from a series of color measurements on the eye. These three functions remain indeterminate, however, with respect to independent linear transformations among them, and hence it is not possible to attribute unique spectral sensitivities to a supposed set of three different light sensitive structures in the eye. If three such separate mechanisms do in fact exist, as in accordance with the conventional Young-Helmholz trichromatic theory of color vision, it seems highly probable that their geometrical distribution in the retina, and perhaps even their size, will differ, and so therefore must their relative resolving power, and from the chemical differences in their manner of light response, it seems even more probable that their temporal characteristics will differ. Accordingly, it should be possible by a careful examination of the properties of spatial resolving power, time sequential color mixing and others, in their dependence on wavelength, to shed more light on this tristimulus theory and perhaps determine uniquely the spectral sensitivities of the three receptors. For a conjectural example of this type of investigation, consider a series of experiments in which the eye's spatial resolving power is related to the intensity and wavelength of the light used. Now each of the three types of receptors should have its own characteristic in this regard and the intensity versus resolution curve of each for a given wavelength should be in inverse proportion to the spectral sensitivity of that type. It may even be that with this sensitivity discounted, this characteristic of the three types is otherwise the same. Of course, the data as obtained do not represent directly these separated curves, but rather the combined resolving power. Making the mild assumption, however, that all three contribute to the resolving power in the same sense, and that the action of the receptors is to some extent independent, one should be able to deduce the individual curves, especially if the data are of a sufficiently precise and critical nature. A similar course of investigation is open in regard to the temporal resolving power of the different color receptors, such as is evidenced in flicker perception and in the merging of colors that are sequentially viewed.

It is thus to be expected that by the application and extension of the methods of colorimetry to the more general case of space and time varying light intensity functions, perhaps in the manner above indicated if necessary, the dependence of the visual characteristics on the spectral properties of such light will be reduced to a dependence on a three parameter set of tristimulus values which themselves are functions of position and time. Once again the engineer would find it convenient if visual resolving power and like properties were expressed in direct relation to the functional characteristics of the Fourier expansions of these tristimulus functions, and it would seem that some means of connection could be developed. For instance, consider a pattern of alternate light and dark gray bars of equal width which has been reduced in size to the point where the division of the area into bars is just barely perceptible by the eye. Now the light intensity in its dependence on the lateral coordinate is a square wave, but since the grossest detail of the area is just barely perceptible, the eye must almost surely be indifferent to the sharpness of the corners of the square wave, the steepness of the sides, and similar smaller details. Therefore one may be persuaded that the only parameters in the light intensity pertinent to the resolution in this case are those connected with the average intensity and the fundamental term in the Fourier expansion, or perhaps better still, connected with some mean value of the intensity gradient. Whatever the form of these connections, however, there can be no doubt of the engineering value of their formulation.

Colorimetry

Within the broad range of light intensities to which the eye is ordinarily subjected, and excepting the two extremes of nearly complete darkness and of viewing highly incandescent bodies, it has been found quite generally that a substantial portion of the properties of most people's eyes which have to do with the perception of color in foveal vision can be tabulated and codified under the comparatively simple system known as trichromatic colorimetry. While the colorimetric properties of different eyes are different, and even the properties of a given individual will change for various reasons and from time to time, it has been found that the different sets of characteristics are, in most cases, sufficiently closely grouped about a mean, that the latter, expressed as the characteristics of a standard observer, is satisfactorily representative of the group, and that in the cases where the differences in properties are considerable, the departures are so marked and of such a nature as to be clearly the result of inherent abnormalities which are usually referred to as forms of color-blindness. Thus in the discussion which follows, reference to the characteristics of the standard observer may just as well be interpreted as reference to the characteristics of any individual normal observer, excepting for small quantitative differences.

Largely as a consequence of the tristimulus system of codification found possible in colorimetry, there has been developed the theory of three color vision in which the eye is thought to contain receptors possessing three different spectral sensitivities. These three receptor types are all associated with the cones in the retina, for it is the cones which are believed responsible for color vision. The rods, also light-sensitive elements in the retina, are responsible primarily for scotopic or darkness vision, and in any case are believed not present in the foveal part of the retina. Colorimetry itself, however, can be understood entirely as a phenomenological description of the color properties of the eye, and is in no way dependent on any explicit theory of vision. The present discussion is from this phenomenological standpoint.

Colorimetry is concerned with those of the visual properties directly associated with the wavelength dependence of the intensity of the light being seen, that is, with purely color perception, and strictly speaking, is confined further to the correlation and codification of color matches and color mixtures. It is concerned with perceptions of solely spectral origin, although in general all of the properties of the eye are affected in more or less degree by this intensity dependence on wavelength, and conversely the spatial and temporal dependence of the intensity have their effects on color perception. For example, visual acuity depends upon the colors and color relationships in the field of view, while in the opposite sense, the color perception of one region of the field of view may be altered by the properties of the immediately adjacent areas. Colorimetry seeks to avoid complications of this nature and in the collection of data, as in its principle instrument which is the colorimeter, time variations and spatial inhomogeneities are carefully suppressed, with perhaps the exception of one bifurcation of the field of view to enable color matches to be made.

The colorimeter consists essentially in simple optical means for presenting to the eye two immediately adjacent, uniformly luminous fields, usually two adjacent semi-circular discs, or possibly two concentric discs, one an annular ring surrounding the other, and means for quantitatively altering the light intensities and combinations in each of the two fields separately. One of these fields, the sample field, may take a variety of forms, depending on the kind of sample that is to be examined. For instance, it may be a pigmented surface illuminated with a definite light source, the combination of pigment plus illuminant constituting the sample as a whole since the kind of light emitted by the field is dependent on both. The basic colorimetric data, however, are concerned only with the visual properties of the light itself, and not with the origin of its spectral characteristics, and, as will appear presently, it is better that the sample distribution come directly from a source and illuminate a white field in the colorimeter so that other lights can be mixed in with it. Thus the sample field and the standard or matching field of the colorimeter may be white matte surfaces, capable of being illuminated simultaneously by several light sources whose intensities are continuously variable while their respective relative spectral distributions remain unchanged. This variation of the intensity of a source may be accomplished by the suitable interposition of a variable aperture in the optical path from the light generating mechanism itself to the reflecting surface, or perhaps by providing a variable distance so as to utilize the change in illumination of the surface with its distance from the light generator, the dimensions of the emitting surface of the latter being small relative to the distance. Whatever the method, care must be taken to preserve the spectral character of the resultant light appearing in the colorimeter. A white surface is chosen for this same purpose. Its reflectance must be independent of the wavelength of the light. Preferably the reflectance should also be perfectly diffuse and follow the Lambert cosine law. That is, the surface should be matte, in order to eliminate insofar as possible complications arising through the fact that the surfaces are illuminated by the different sources in different directions, and that then too, the viewing direction may be not always exactly normal to the surface.

Measurements with the colorimeter are performed by adjusting the intensities and mixtures on the standard side until the line of demarcation between the sample field and the standard field disappears. There is thus no visual distinction between one side and the other (except location, of course) and the two colors are said to match. Sometimes it is advisable to rotate the demarcation line by some optical means, so as to allow for inhomogeneities in the retinal structure and perhaps for astigmatism, but in general it is found that there is a small interval in the intensity settings over which the match is equally effective. This is an indication of the limited color discriminating capabilities of the eye. Furthermore, with the same adjustment of the instrument, a mismatch may be observed at some later time or by a different observer. This is indicative of the variability of the visual characteristics, as mentioned before, but nevertheless on a statistical basis a match can be regarded as an observable relationship.

Following the above outline, two different forms of conventional colorimetric data are obtained in the colorimeter, depending upon whether the light sources for the standard or matching field consist of three primary sources, as defined below, each with a constant relative spectral distribution, or whether they consist of one source of white light and a monochromatic source of variable wavelength. Excepting for this last named source, throughout the discussion it is meant that relative spectral characteristics are always unaltered. A white light commonly used is the so-called "C" illuminant which has a certain standardized spectral characteristic and is rather easily reproduced. As will be seen later, there is in principle no restriction on the spectral nature of the white light, nor on the three primary sources, for data taken with them can be, from their spectral distributions, converted over to data corresponding to any other set of sources. Of course, whenever the color gamut of the samples to be examined will allow, it is more convenient and more precise if the matching sources are capable of effecting direct matches. In general, directly additive matches, in which the matching sources illuminate the matching field alone, and the sample source illuminates the sample field alone, are not always possible. In order to effect a match it is sometimes necessary to divert one or more of the matching sources from the matching field to the sample field, and, so to speak, dilute the light from the sample source with the lights from some of the matching sources. In these cases the matching light intensities added to the sample are said to be added to the matching field with negative intensities. With allowance for the use of such a device, which will henceforth always be assumed, it is possible to match any sample with light from any set of three primary sources or with light from the white and monochromatic sources.

Two color sources are said to be primary with respect to each other if, when they are regarded in the two fields of the colorimeter no non-zero adjustment of their intensities can produce a match; three color sources are said to be primary with respect to each other, and form a set of primaries, if no one can be matched by any combination of the other two.

The results of colorimetric investigations with different spectral distributions and also with monochromatic light of various wavelengths have led to the formulation of two fundamental laws which express the trichromatic basis of color representation and the linearly additive colorimetric nature of light:

1. ***The light from any sample light source, at any given intensity, which is in its effect a color stimulus or simply a color, can be matched by one and only one combination of a given set of three primary sources.***

2. ***The combination of two colors matches with the combination of their matches.***

In addition to these laws, from which follow the relations that will be used in the subsequent discussion, the colorimetric data yield three functions or combinations thereof, which are conventionally tabulated as the colorimetric distribution functions and often symbolized by $\bar{x}$, $\bar{y}$, and $\bar{z}$, numerical quantities dependent on wavelength. These functions are to be interpreted as the intensities of the three ideal primary sources needed to match unit intensity monochromatic light of the corresponding wavelength. It is important to realize that, although functions of wavelength and at the same time representing intensities, these quantities by no means represent the spectral intensity distributions of the primary sources. Their connection with the latter is quite indirect. Furthermore, when the data are converted to a given set of primaries in actual use, the resulting functions likewise do not describe the spectral distributions of the primaries. They state what intensities of the primaries, with all their spectral distributions, are required to match various monochromatic colors. These colorimetric functions state, therefore, what total intensities of the primaries are required in a given case, not what spectral distributions in the primaries are required. Thus it turns out that while a chosen set of primaries may have zero intensities over considerable portions of their spectra, the appropriate distribution functions of wavelength are never zero (within the visible spectrum) except perhaps at a few singular wavelengths.

The various colorimetric relationships consequential to the two basic laws and necessary for the proper utilization of the numerical data are usually expressed in a rather cumbersome notation which requires the frequent writing of triplets of symbols and formulas of similar type. Since these relationships are given here in some detail, a more convenient notation has been adopted. Fortunately this notation is not new. It has been taken from the absolute calculus and should be quite familiar to those acquainted with tensor analysis. It will be described here from the point of view of its special application.

The first colorimetric relationship is an obvious deduction from law 2, and may be stated as follows:

Relation 1—If two colors match, and their intensities are changed by the same factor, the resulting colors will also match.

Accordingly, it is convenient to write an appropriate factor explicitly. Let S be a sample source which is being matched in the colorimeter, as by a set A of three primary sources, A_1, A_2, and A_3. If E is the total energy intensity of the sample as it illuminates the field of the colorimeter, let $E\,I_s^{A_1}$ be the corresponding intensity of the A_1 primary source as it effects the match with S. Then $I_s^{A_1}$ is the relative intensity of the A_1 primary, or the actual intensity if the sample had unit intensity. The superscript identifies the primary source effecting the match and the subscript identifies the sample source matched. If the latter is monochromatic, of wavelength λ, it is more convenient to write the relative intensity as $I_\lambda^{A_1}$. For the given set of primaries, the three quantities $I_\lambda^{A_1}$, $I_\lambda^{A_2}$, $I_\lambda^{A_3}$, depending on the wavelength of the monochromatic sample, are thus functions of λ. To have the notation consistent, since the spectral intensity distributions of the various sources are also functions of wavelength, they will also be represented by letters with λ as an indice. Again it is convenient to separate the total intensity and the relative spectral distribution, so the spectral intensity distribution of the sample S is given by $E\,D_s^\lambda$, while the relative spectral distribution of the primary source A_1, or its actual intensity distribution when adjusted to give unit total intensity, is given by $D_{A_1}^\lambda$. λ is written here as a superscript and when in connection with I, as a subscript, in part to show the different dimensionality of the two quantities.

Consider now sets of three primary sources A_1, A_2, A_3, and B_1, B_2, B_3, etc. By a, or by a', is denoted any one of the sources A_1, A_2, A_3, and by b, or b', any one of B_1, B_2, B_3. Thus any relation involving a is to be read first with $a = A_1$, then with $a = A_2$, then with $a = A_3$, and so on. Thus I_s^a stands indifferently for any of the three relative intensities $I_s^{A_1}$, $I_s^{A_2}$, $I_s^{A_3}$, and any equation written in terms of a holds whether a is replaced by A_1 (the same throughout the equation), by A_2, or by A_3. Similarly, when λ appears as an indice, the equation holds for all values of wavelength. In this manner there are nine quantities I_b^a denoting the three sets of each three relative intensities of the A primary sources matching the B sources and there are nine different quantities I_a^b for the relative intensities of the B primary sources matching the three A sources.

The usual summation convention is also adopted for this present notation: whenever, in a product, the same indice appears twice, the expression is read as the sum of the several like expressions in which this indice is replaced successively by each of the values it can assume. For example:

$$E\, D_a^{\lambda}\, I_s^a \text{ is read as } E\,(D_{A_1}^{\lambda}\, I_s^{A_1} + D_{A_2}^{\lambda}\, I_s^{A_2} + D_{A_3}^{\lambda}\, I_s^{A_3})$$

Thus there are 81 quantities $I_b^a\, I_{a'}^{b'}$ (with regard for the summation convention the primed and unprimed indices are not considered the same, even though they refer to the same primary set) 9 quantities $I_b^a\, I_{a'}^{b}$, and 1 quantity $I_b^a\, I_a^b$. In the event that λ is a repeated indice, since λ can assume a continuum of values, the notion of summation is replaced by that of integration with respect to wavelength. For colorimetric purposes it is sufficient that the limits of integration include all visible wavelengths, but the original notion of total energy intensity must likewise be confined to the same limits, about 380 to 750 $m\mu$[3]. In dealing with photoelectric devices in connection with colorimetry it is therefore most essential to ensure that the physical apparatus is also thus confined. From the definition of D_s^{λ} as the relative spectral distribution, it follows accordingly that

$\int_{380\, m\mu}^{750\, m\mu} D_s^{\lambda}\, d\lambda = 1$ for any source S. From the present convention a similar integral is formed by the repetition of the indice λ, for:

$$I_{\lambda}^a\, D_s^{\lambda} \text{ is read as } \int_{380\, m\mu}^{750\, m\mu} I_{\lambda}^a\, D_s^{\lambda}\, d\lambda$$

Relation 2—Since any sample color of relative spectral distribution D_s^{λ} may be regarded as the combination of all the monochromatic colors with appropriate relative intensities in that distribution, while the matches to the latter are given by the three functions I_{λ}^a the combination of these matches is the match to the original color, by application of law 2. In terms of the adopted notation, this colorimetric relationship may be expressed as: $I_s^a = I_{\lambda}^a\, D_s^{\lambda}$.

Should members of a primary set B be employed as samples, this expression would be written as: $I_b^a = I_{\lambda}^a\, D_b^{\lambda}$.

[3] $m\mu$ = millimicron.

Relation 3—If a sample source S at intensity E is matched by a set of three A primary sources with intensities EI_s^a, and if each of these primaries, when adjusted to unit intensity, is separately matched by a second primary set B with relative intensities I_a^b, then the sample S can be matched directly by the B set with intensities EI_s^b where $I_s^b = I_a^b I_s^a$. Replacing S by λ for monochromatic samples, one finds also that $I_\lambda^b = I_a^b I_\lambda^a$.

Relation 4—A consequence of the above relations is that if each of primary set A with unit intensity is matched by a primary set B with relative intensities I_a^b and if each of the set B is matched by the primary set A with relative intensities I_b^a, the two sets of intensities are related by the equation $I_b^a I_{a'}^b = \delta_{a'}^a$
where the symbol $\delta_{a'}^a$, the Kronecker delta, is 1 or 0, depending on whether $a = a'$ or $a \neq a'$, respectively. The three systems of three simultaneous linear equations can be solved either for the I_b^a or the I_a^b. For example:

$$I_{A_1}^{B_1} = \frac{I_{B_2}^{A_2} I_{B_3}^{A_3} - I_{B_3}^{A_2} I_{B_2}^{A_3}}{| I_b^a |}$$

where, in the determinant, the superscript denotes the row, the subscript the column.

Relation 5—It follows either from relation 4 or more directly from the uniqueness rule expressed in law 1 that if two identical primary sets A are matched with each other, $I_{a'}^a = \delta_{a'}^a$

Relation 6—The quantities $I_{a'}^a$ in relation 5 may be computed with the aid of relation 2, and take the form: $I_{a'}^a = I_\lambda^a D_{a'}^\lambda = \delta_{a'}^a$

Relation 7—The condition that a set B of three sources be primary, that no one can be matched by a combination of the other two, can of course be applied to the matches by set A of each of the set B. Thus $I_{B_1}^a$ must not be a linear combination of $I_{B_2}^a$ and $I_{B_3}^a$, that is, that there be no non-zero solutions x^b to the three simultaneous equations $I_b^a x^b = 0$.

This requires that the determinant of the coefficient of x^b be other than zero: $| I_b^a | \neq 0$.

Relation 8—The luminosity of a color is not defined solely in terms of colorimetric matches, but some of the most important properties of the well-known luminosity function or curve of eye sensitivity versus wavelength are contained inherently in the standard colorimetric data.

This is by virtue of the properties believed of the luminosity of a color, that it is a single numerical quantity, the same for all colorimetric matches, and that the luminosity of a combination of colors is the sum of the luminosities of the components. The luminosity L_λ of unit intensity monochromatic colors is in consequence a linear combination of the functions $I_\lambda{}^a$ where the A sources are any primary set. This can easily be seen if the luminosity of a monochromatic color is expressed in terms of the luminosities of the matching primary set which in turn are the sums or integrals of the luminosities of their monochromatic components: $L_\lambda = I_\lambda{}^a D_a{}^{\lambda'} L_{\lambda'}$.

The expression on the right is obviously a linear combination of the three functions $I_\lambda{}^a$. L_λ is also known as the standard visibility function.

Luminosity is one example of the several visual perceptions connected with the spectral distributions of the light received yet recognized as common to, or possessed in different degree by different distributions which are not colorimetric matches. Another such perception is that of hue, another, that of chroma or saturation. They are properties of color possessing a certain degree of uniqueness with respect to other properties and yet not always capable of objective operational definition. For instance, the ability of hue identification and discrimination differs very greatly among individuals and at the same time is very much a matter of experience and training. Conceivably it may be an entirely acquired visual ability, learned from ordinary experience with the dilution of dyes and pigments by white. Of these visual perceptions, luminosity appears to be the best defined and established. Large luminosity differences between colors are easily recognized, even though the colors are otherwise very different. When the relative spectral distribution is unchanged, luminosity varies directly with intensity and this as a simple proportionality. There is of course substantiating evidence for this assumption. For example, the luminosity, measured by comparison with a steady source of the same relative spectral distribution, of light subject to rapid time variations, is the same as the time average of the instantaneous luminosities. The homochromatic comparison of the luminosities of the light from two fields may be done very easily in a colorimeter or similar instrument, which is then regarded simply as a photometer. Heterochromatic photometry is substantially more difficult, and since exact comparison when the colorimetric properties of the sources are widely different is almost impossible, the measurements are usually done in a step by step process in which a series of slightly different sources are compared. Thus in the determination of the luminosity vs. wavelength function, two monochromatic lights of almost the same wavelength are equated in luminosity by adjusting their relative intensities until their visual differences in the colorimeter appear to be minimized. Then this procedure is repeated, using one of the monochromatic lights already used, and so on down the wavelength scale.

Although the procedure of step by step comparison outlined above constitutes the necessary operational definition of luminosity, from which it is subsequently found that the luminosity function is a colorimetric function in accordance with relation 8, a very much less tedious method of heterochromatic photometry is available in the use of the flicker photometer. Here the two light sources to be compared are alternatively presented in the same field of view and their intensities are adjusted so as to minimize the critical flicker frequency. This is the frequency of alternation of the field at which, as the frequency increases, the flicker arising from the alternation just disappears. While this method is subject to the criticism that flicker perception is a separate phenomenon, connected primarily with time variations and not necessarily at all with luminosity perception, and furthermore, flicker perception is due, at least in part, to the rods in the retina, which are perhaps inactive in color vision, nevertheless, the flicker photometer has been considered by many as giving good results in heterochromatic photometry.

With the aid of the colorimetric laws and relations which have been given, all that is necessary for the codification of colors for reproducing and matching is the numerical values of the functions I_λ^a for some definitely specified set of primaries A which are accepted as standard. In accordance with relation 2, therefore, all spectral intensity distributions D_s^λ would be uniquely characterized in their colorimetric properties by the appropriate sets of three quantities I_s^a. Since both the functions I_λ^a and D_s^λ are empirical in origin they are not conveniently represented by simple functions and the process of computing the integrals $I_\lambda^a D_s^\lambda$ is usually one of numerical integration. For this procedure it is very inconvenient to have negative values of I_λ^a and since the choice of the standard primary set is arbitrary in the first place, it would be best if a choice could be made which would, if possible, avoid these negative values. Furthermore, the luminosity function is a linear combination of the general expressions I_λ^a and it is also always positive. Hence the total luminosity can be a very convenient extra dividend obtained when computing the I_s^a values of a color, if the luminosity function is itself one of the I_λ^a. Now supposing the actual color data to be taken with a definite set of primaries B, by relation 3, $I_\lambda^a = I_b^a I_\lambda^b$. The quantities I_b^a are determined uniquely from the quantities I_a^b by the equations of relation 4: $I_b^a \, I_{a'}^b = \delta_{a'}^a$. The quantities I_a^b, on the other hand, are fixed by relation 2, $I_a^b = I_\lambda^b \, D_a^\lambda$. If the functions D_a^λ are chosen arbitrarily it is possible to assign any set of values to the quantities I_b^a. Thus it is at once possible to choose I_b^a so that $I_\lambda^{A_2}$ is the luminosity function. Since the latter is always greater than zero for visible light, as has already been pointed out, two more independent combinations of the I_λ^b can be found which are non-negative. Of course the values at either end of the visible spectrum are all always approaching zero, and thus in particular it is possible to find at least one such independent combination[4] which approaches zero somewhere toward the middle of the visible range. In addition, the function I_λ^a may be multiplied by appropriate numerical factors so that all three $\int_{380\,m\mu}^{750\,m\mu} I_\lambda^a \, d\lambda$ have the same value. Hence, a light source whose intensity is independent of wavelength, called the standard "E" illuminant, will have tristimulus values all of the same numerical value. It so happens in the colorimetric data that for wavelengths greater than about 550 mμ, one of the functions, say $I_\lambda^{B_3}$, can be expressed as a linear combination of the other two, to a very good approximation. It is therefore convenient to choose the I_λ^a so that the residual third function is approximately zero in this wavelength region.

The above considerations have been taken into account and the choice of the standard distribution coefficients I_λ^a has then been fixed[5] by the specification of the trichromatic coefficients of a certain three monochromatic wavelengths and the standard "B" illuminant, which has a specified spectral distribution that, incidentally, gives a white color. The conventional designation for the distribution coefficients, which have been standardized and tabulated is, $\bar{x}$, $\bar{y}$, and $\bar{z}$, the $\bar{y}$ function (of wavelength) being also the luminosity function. (Figure 1.) It is not possible to state exactly what the spectral intensity distributions of the three corresponding ideal primaries are because the distributions are not uniquely specified nor required by the functions I_λ^a or $\bar{x}$, $\bar{y}$, $\bar{z}$. Relation 6, that $I_\lambda^a \, D_{a'}^\lambda = \delta_{a'}^a$, however, does give evidence of one important condition on the D_a^λ, namely, that for all the I_λ^a to be positive, for some wavelengths the D_a^λ must be negative. Now since

[4] In the neighborhood of 503 mμ the $\bar{x}$ function comes near but does not quite touch zero. This facilitates computation and saves coordinate space in the chromaticity diagram by bringing the curve of spectral colors close to the y axis.

[5] The transformation from existing experimentally-derived distribution coefficients to the standard distribution coefficients was thus specified by the Colorimetry Committee of the Eighth Session of the International Commission on Illumination, Cambridge, England (September, 1931).

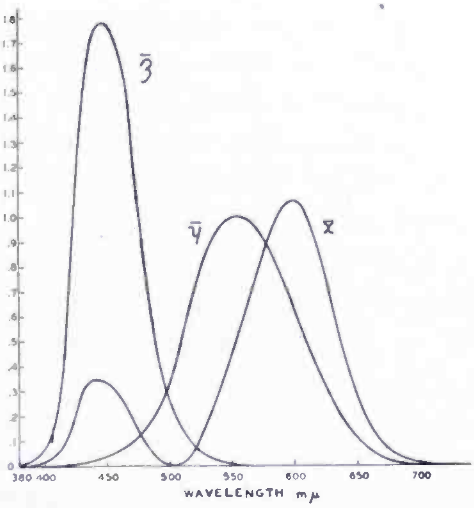


Fig. 1—Standard tristimulus values of equal energies of the spectral colors.

no light source can have negative intensity, over any part of its spectral distribution, it is clear that the ideal primary sources are imaginary. To be sure, they can be obtained directly in the colorimeter by the artifice of negative addition but each primary would then con-

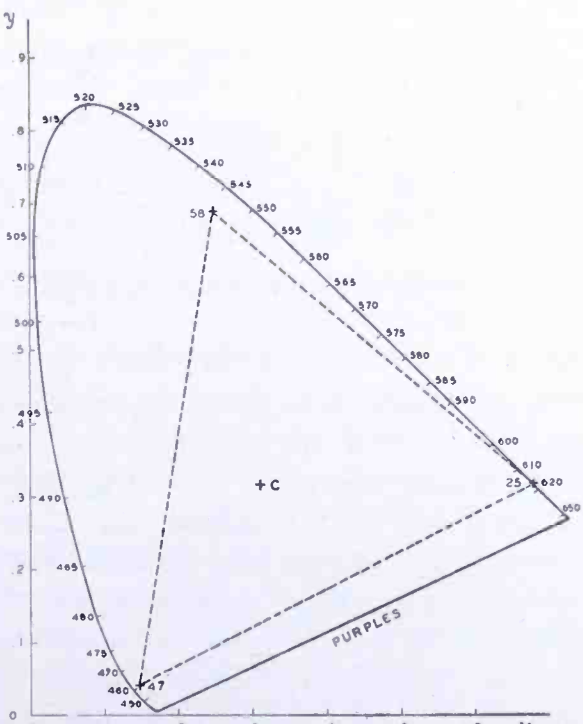


Fig. 2—I.C.I. standard chromaticity diagram, showing the line of spectral colors and of pure purples, the location of the standard "C" illuminant, and No's. 25, 47 and 58 Wratten filters in combination with that illuminant to illustrate the reproducible triangle of a receiver with these as primaries.

sist of two light sources. Fortunately there is no real need for the direct use of these imaginary, ideal primaries. Their utility arises in colorimetric tables, charts and computations.

In order to secure a graphical representation of the colorimetric relations and of various color differences and similarities, a color chart in the form of a chromaticity diagram (Figure 2) is generally used. While many types of color charts have been invented with this object in view, the standard chromaticity diagram is capable of a fairly complete, objective representation although it cannot be used for direct visual comparison with sample colors unless it be printed with an appropriately shaded colored background. It is constructed from the tristimulus values of the sample colors to be represented on it, and since the diagram is two-dimensional it is appropriate to discount from the color specification that factor having to do primarily with intensity and preserve in the representation those quantities representing properties seemingly more intrinsic to the notion of color. In this way the tristimulus values of a sample, which are the quantities $EI_s{}^a = EI_\lambda{}^a D_s{}^\lambda$ when the $I_\lambda{}^a$ are $\bar{x}$, $\bar{y}$, $\bar{z}$, and are denoted conventionally by X, Y, Z, are formed into two ratios $\dfrac{X}{X+Y+Z} = \dfrac{I_s{}^{A_1}}{\sum_a I_s{}^a}$, $\dfrac{Y}{X+Y+Z} = \dfrac{I_s{}^{A_2}}{\sum_a I_s{}^a}$. These two ratios, called the trichromatic coordinates of the sample, x and y, have values between 0 and 1, and together with the value of Y, the luminosity, completely specify the colorimetric properties of the sample in the same way as do the quantities $I_s{}^a$. There is a third trichromatic coordinate, $z = \dfrac{Z}{X+Y+Z}$, but since, always, $z = 1 - (x + y)$, it can usually be ignored.

The chromaticity diagram is based on Cartesian coordinates with respect to which are plotted the trichromatic coordinates of the samples to be represented, the value of x along the x axis and the value of y along the y axis. In this manner, all colors which differ in visual characteristics only in luminosity, correspond to the same point on the chromaticity diagram. Every monochromatic line of the visible spectrum has such a point and the points corresponding to successive wavelengths turn out to be on a horseshoe shaped curve. The red side of this horseshoe is very nearly a straight line, as is to be expected when, in this wavelength region, only two of the functions $I_\lambda{}^b$ are strongly independent.

Relation 9—One of the most important characteristics of the chromaticity diagram, which is a simple consequence of its method of construction, is that the point standing for the additive combination of two colors lies at the center of gravity of the two points standing for each of them, they being assigned weights proportional to their respective quantities $\sum_{a} I_s^a$ or $X + Y + Z$. This notion similarly applies to the combination of several colors.

It therefore follows that all colors which are real, meaning that the spectral distributions of the light of the colors are made up of positive intensity contributions of the various monochromatic wavelengths, must lie inside of the area bounded by the monochrome horseshoe and the line joining its two ends. Daylight and sunlight, which each have roughly uniform intensities of light throughout the visible spectrum, are in this manner represented by points lying near the center of the horseshoe.

Relation 10—In consequence of the center of gravity relation 9, it is clear that the points representing a set of primary colors on a chromaticity diagram are necessarily non-colinear. Any color matched by a positive combination of such a primary set, whatever the relative intensities, must lie on or within the triangle formed by the three points representing the primaries.

Some additional properties of the chromaticity diagram may be noted as follows. The purity of color, corresponding very closely to the visual perception of saturation or chroma, is intended to indicate how nearly monochromatic a color is, or conversely, how degraded with white. If a straight line is drawn from the point on the chromaticity diagram representing white, for which the standard "C" illuminant is usually taken, through the point associated with a given color, on to meet the monochrome line or line of spectral colors, the purity of the given color is defined as the ratio of the line segment between white and the color, to the total length. If the given color is a purple, it is necessary to close the diagram by joining the two ends of the monochrome line with a straight line which may be thought of as representing the pure purples. The meeting point with the monochrome line mentioned above indicates a particular wavelength which is known as the dominant wavelength of the given color sample. In the case of purples this wavelength is not so defined, but rather is taken at the intersection with the monochrome line of the line from the given color extended through white. This corresponds, as will be seen later, to the dominant wavelength of the color complementary to the given purple, and therefore is distinguished by the suffix *c*.

From the properties of the diagram already given, it can be seen that the combination of white and monochromatic light of the dominant wavelength, or its equivalent negative combination of white and the complementary dominant wavelength in the case of purples, when in amounts indicated by the purity and the total luminosity, will provide an accurate match of the given color. Obviously therefore, the location of a color on the chromaticity diagram is equally well specified either by its trichromatic coordinates or its dominant wavelength and purity.

In connection with judgment as to the approximate reproduction of colors, a very useful representation of color tolerances may be made on the chromaticity diagram by plotting local contours of least chromaticity differences perceptible to the normal eye.

The combination of colors as it is understood in colorimetric relationships and data and as it is expressed on chromaticity diagrams, involves the combination of light fluxes impinging on the eye and has nothing whatever to say about the manner in which the spectral intensity distributions were secured originally. The artifice of negative addition of light has referred merely to the addition of light flux to the sample side of the colorimetric balance, rather than to the matching side. Unfortunately, from the point of view of terminology, the notions of primary colors, color combinations and charts have been applied to a large group of processes associated with color production which themselves have received a collective misnomer, subtractive processes. They are concerned with the ways in which dyes and pigments modify the light impinging on them and the consequences of mixing and combining different pigments and dyes. For the present brief discussion of subtractive processes, the actual atomic properties which give rise to the color characteristics of these substances will not be considered but the spectral reflectances and transmittances will be already assumed.

From the point of view of colorimetry, pigments possess, at least to some extent, a very different intrinsic property from that of dyes, insofar as the solid flakes or particles of pigment are individually opaque to the incident light. If this opacity is complete, each tiny ray of incident light is acted upon by only one kind of pigment in a mixture, and the reflected light which is viewed is really the additive combination of rays from the different kinds of pigment. A given kind of pigment selectively absorbs a greater part of the incident light of some wavelengths and reflects a greater part of other wavelengths and in this sense some of the light incident on the material has been subtracted in the reflected beam. However, the amount is proportional to the amount incident so that the process can much more accurately be called division, or alternatively, logarithmic subtraction.

The spectral quality of the light reflected from a pigment, and hence in general its apparent color properties, depends as much on the quality of the incident light as it does on the reflectance characteristic of the pigment. It is therefore meaningless to speak of the colorimetric quantities associated with a pigment, and to plot it on a chromaticity diagram, unless its associated illuminant is also specified, either explicitly or as understood. Generally, it is desirable that the illuminant be near to white with a fairly uniform spectral distribution. For this purpose, some standard illuminant, as the "C" illuminant, is often used. The latter has a definite spectral characteristic resembling daylight and can be reproduced easily.

Once the illuminant is fixed, so that any given pigment may be identified with certain tristimulus values, so long as the opacity condition on the particles is satisfied, mixtures of pigments behave colorimetrically just as mixtures of light fluxes, and one can obtain primary sets of pigments which yield in additive mixtures precisely similar results to primary light sources, except that the intensity of the reflected light is controlled by the illuminant. On the other hand pigment particles seldom really satisfy the opacity condition, or the structure of the pigment surface may be such that light rays are reflected from one grain to another before leaving the surface. In such cases the pigments tend to assume the characteristics of dyes, where the additive color mixture properties are scarcely ever obtained. It is then too that the spectral reflectances depend to some extent on angles of incidence and reflection.

The spectral quality of light incident on a dyed surface is affected in much the same way as occurs in the passage of light through a dyed, transparent medium. The incident light passes through the dye, is reflected by the underlying surface, most often of approximately uniform reflectance, that is, nearly white, and again passes through the dye on its way out. Thus the spectral quality is altered by the selective absorption of some wavelengths of the light passing through the dyed medium. The characteristic transmittance of the dye, like the reflectance of a pigment, is a proportional function, and while the light which is absorbed by the dye may be thought of as subtracted, since the amount is in proportion to the amount present the subtraction is really logarithmic. When two dyes are mixed, the same ray of light is acted on by both of them and the resultant transmittance is the product of the separate transmittances of the two. Thus, even if the illuminant is specified and the light passing through a definite quantity of a given dye therefore may be assigned a unique set of tristimulus values, the light which passes through a combination of two dyes is in general totally different from an additive combination of the light passing through each of the dyes separately. In fact it is not possible to choose dyes at all which possess this additive property for more than one set of densities. Thus, there is no unique relation between the colorimetric specification of the original dyes and of their combination. Their mixture characteristics depend entirely on their original transmittances, are colorimetrically non-linear, and do not obey the colorimetric relations for the combining of colors. For example, there is no single point on the chromaticity diagram which corresponds to a given dye, even with a specified light source. The light transmitted by zero density of dye has, of course, the trichromatic coordinates of the source, that is, in the white, and as the density of the dye increases, the point moves outward along some usually curved path. With the density increasing without limit, the point approaches the position of the monochromatic light of wavelength corresponding to the peak transmission of the dye, although the total amount of light transmitted becomes vanishingly small. Another example of the irregularities of the so-called subtractive systems is encountered when mixing a dye with each of two dyes which at the given density have identical appearances. Notwithstanding the colorimetric equivalence of the latter two dyes in combination with the chosen white source, the final mixtures may yield entirely different colors.

The identification of primary dyes is substantially more difficult than the identification of primary light sources, for while a sort of condition of being primary may be imposed, that the transmittances at all finite densities of each be unobtainable by any combination of the other two, this condition is not sufficient to insure that all colors can be matched, even allowing for control of the intensity by means of the light source. Furthermore, there is in subtractive systems no useful analog to the negative combinations in the colorimeter of the additive color systems. This leads to the fact that the notion of primary, in connection with subtractive processes, is, besides being less precise, more restricted than has been used here for the additive combinations. Thus, through the artifice of negative addition, an unlimited number of primary sets were possible in the additive combinations even though relatively few could match a comprehensive gamut of colors by positive addition alone. An example of such a preferred set of primaries is a high purity red, a high purity green and a high purity blue. In subtractive processes, since the same gamut is to be covered, the additive primaries themselves must be attainable. Finally, for reproduction purposes some simple law of combination, such as the linearity of the additive systems, must be obtained on practical grounds. Ideal dyes may be conceived of which would achieve this, and accordingly are identified with the manner in which they control the effective amounts of appearance of the additive primaries. Thus with white light at the start, the dye which controls the amount of blue is the one which removes blue from the white light, hence "minus blue". Applying the rules of additive mixtures to white light and a negative amount of blue, we see from the chromaticity diagram the resulting color is yellow. Hence a yellow dye which has that color by virtue of fairly uniform high transmission in the red and green parts of the spectrum and sharp absorption in the blue is capable of being the "minus blue" subtractive primary. Similarly the "minus red" is a blue green or cyan dye, and the "minus green" is a magenta or purple, which transmits red and blue light and absorbs green. In most cases pigments partake of both subtractive and additive properties because of partial reflection and partial transmission of light by the particles. Thus adding a "minus red" pigment also increases the "plus blue". In consequence of this and the fact that the subtractive characteristics rarely approach the ideal of sharp spectral selectivity, the common primaries in pigments are generally blue, yellow and red.

Now the color produced by a "minus" dye in combination with white light, and the color controlled by this dye in subtractive processes, for instance "minus blue" or yellow and blue, bear a special relationship to each other which is often of particular artistic interest. The one color is the complement of the other. The additive combination of the two produces white. They appear conjointly in the well-known eye fatigue phenomenon in which a bright design of one color is viewed fixedly against a white background, and then a blank white surface is viewed, wherein the same design in the complementary color appears for a time as an illusion. As has been indicated before, what is called white requires an arbitrary standardization, but all the properties of complementary colors are consistent if the same white is referred to. Yet while the location of the white fixes the line on the chromaticity diagram on which the complement of a color occurs, the degree of saturation of the complement is not specified. If the connection of complementary colors through the subtractive process were taken as the unique definition it would follow that a color of high saturation would have as complement a color of low saturation, and so on. On the other hand, a requirement that a color and its complement have the same saturation would be equally unique in specification, while the condition that the additive sum be white would still be met but with different relative luminosities. Such a definition seems to be more compatible with the artistic ideas of complementarity. Sometimes it is not necessary to define exact correspondence between complements, but only between complementary hues, but this too is not free of difficulty for the dominant wavelength of a hue depends to a slight extent on the saturation.

It must be remembered, however, that all true colorimetric specifications are referred to additive combinations of light sources, and that the conditions on additive combinations must always be satisfied to achieve color matches, whether the intermediary steps by which the colors are produced are additive or subtractive. The simultaneous television receiver is an excellent example of the additive method of forming color with superposed light from three different light sources.

Color Reproduction

From the foregoing section on colorimetric matching it should be clear that the visual identity of colors may be stated uniquely in terms of the identity of their tristimulus values. Therefore the function of a color reproduction system is to provide colors with the same tristimulus values as occur in the original. While reproducing the original spectral distributions would indeed satisfy this condition, such a procedure is clearly not necessary, even for exact color reproduction, and it is not possible in a practical sense. Admittedly a three color system which is exactly right for one person may be not quite perfect for someone else, and in addition engineering tolerances may prevent exactly correct reproduction in any case, but in general for any three parameter system, the precision of color reproduction possible on the basis of colorimetric matching is vastly better than can be achieved by any method of approximately reproducing spectral distributions. Therefore, in all systems of color reproduction, in photography, printing, and color television, and wherever a three-color system is used, no attempt is made, nor would it be desirable, to approximate the spectral distribution of the original colors. This is an important example in communications of the selection of physical stimuli of markedly different physical specification from that of the original in order to give rise to the same sense perception.

In terms of the facts and relationships of colorimetry, to achieve accurate color reproduction in a color television system, the correct signals must be expressed in the light intensities of the three color sources in the receiver and this requires, in effect, that the television system operate as an electronic colorimeter, individually for each picture element. Assuming linearity of the three receiver primary light sources, or appropriate correction for non-linearity, the electrical signals controlling these sources must, with two exceptions, correspond, when the transmitting camera is viewing a color sample S, precisely to the quantities EI_s^b which have been discussed in the preceding section and represent the intensities when matching S in the colorimeter of the three primary light sources B which also are the receiver light sources, i.e. kinescope phosphors plus light filters.

The first of these exceptions is merely that in conventional television practice no attempt is made to duplicate in the receiver precisely the same brightness level as occurs in the scene being televised, but the adjustment of scene brightness is left to the selection of the individual viewer. Thus the quantities EI_s^b may be in error by some proportionality factor common throughout the picture.

The second exception arises from the difference in purpose of colorimetric matching and color reproduction. In the former the securing of quantitative relationships was the primary purpose, and the modification of the visual effect of the sample by the negative addition of the primaries was quite admissible. In reproduction the sample is not available to be so modified nor is such desired, for the sample should be reproduced like the original. Therefore complete accuracy of reproduction will be confined to those colors which require only positive combinations of the three primaries. As was shown before, these colors lie on or within the triangle on the chromaticity diagram determined by the three primaries and at the same time their luminosities are limited to the luminosity range of the receiver primaries. Therefore in the design of television receivers it is of considerable importance to select receiver primaries which will cover the desired gamut of accurately reproducible colors. In the present three-kinescope type simultaneous receivers this is very largely a problem in phosphor composition. Fortunately phosphors are available with very satisfactory spectral characteristics although there is ever a need for more luminosity. Some of these phosphors, in combination with optical filters, are capable of saturation so nearly approaching that of monochromatic light, at well-separated points of the chromaticity diagram, that simultaneous receivers can accurately reproduce the colors from virtually all natural and artificial dyes and pigments. They are not able to reach pure monochromatic colors such as a spectroscope will produce but they can easily reproduce rainbow colors which are, as usual, diluted with a small amount of white light. The available color gamut is greater than in most other color reproduction systems and appears adequate.

Within the range of reproducible colors, the function of the camera and transmitting equipment is to derive from the light of the sample S the quantities EI_s^b and impress them on the primary sources at the receiver. The unique physical specification of the sample S is its spectral distribution $ED_s{}^{\lambda}$ which is all the signal information available to the camera. Conventional ideas as to how the camera should obtain the proper signals from this spectral distribution have, until recently, followed the pattern that each of the three signals $EI_s{}^{b}$ should be obtained by a separate operation on the light $ED_s{}^{\lambda}$. Thus in the sequential color camera the plan was for the scene first to be scanned and the "red signal" $EI_s{}^{B_1}$ to be derived for all elements of the picture, then for the scene to be scanned with the "green signal" $EI_s{}^{B_2}$ derived, and so on. In the simultaneous color camera the plan was for all three of these operations to be done at the same time but each by a separate camera tube. That, as they are, these methods are fundamentally unworkable has not always been appreciated, and when it has, the attitude has been largely to do as well as possible with the system and accept the compromise. While not wholly responsible for this unfortunate viewpoint, two erroneous ideas have contributed and it is well to repudiate them explicitly before describing systems which will fully accomplish the desired result.

The first of these misconceptions is a consequence of attempts to approximate the spectral distribution of the original color. This distribution was to be divided into three adjacent blocks, roughly red, green and blue, by being scanned through three filters with rectangular spectral cut-off characteristics. With the three receiver primaries having similar rectangular spectral distributions and being controlled each by one of the signals derived from scanning, the reproduced spectral distribution would approach the original in the manner of a block diagram approaching a continuous curve. If it were otherwise feasible to increase the number of blocks by having perhaps ten or more receiver colors and signal channels, color reproduction by this method would have good quality, but with only three parameters available, the results are necessarily far inferior to colorimetric reproduction.

The second common misconception involves some concession to colorimetry in that it is recognized that spectral sensitivity curves peculiar to the nature of the receiver primaries must be obtained in the color camera. The mistake lies in supposing that the weight by which light of a certain wavelength is to be counted in the camera must be the same as it appears in the receiver primary—in other words, that the contours of the spectral sensitivities of the camera should be the same as the spectral intensity distributions of the receiver colors which they each control. Actually the correct sensitivities are the functions $I_{\lambda}{}^{b}$, which scarcely resemble the primary distributions $D_b{}^\lambda$. If the reasoning of colorimetry is to be abandoned momentarily for the sake of a crude intuitive explanation, the reason that the above idea is wrong can be seen from the fact that a combination of the light from the receiver primaries can give the same visual impression as light whose wavelengths are totally absent from all these primaries.

The fundamental idea as to how the color camera is to derive the three signals $EI_s{}^b$ is based directly on the colorimetric relations by which these same quantities may be computed, namely by a process of multiplication and then integration with respect to wavelength. Thus $EI_s{}^b = ED_s{}^\lambda\, I_\lambda{}^b$, with the integration with respect to wavelength indicated by the repeated indice λ. This process might be performed very easily by a photo-sensitive camera tube, for its total response is the integral of its response at each wavelength, and this spectral

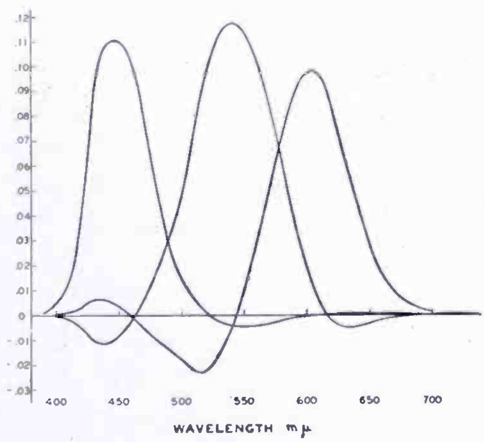


Fig. 3—Transmitter camera sensitivities required for the primaries of Figure 1, assuming that each receiver primary is separately derived. (The curves show negative values for some spectral wavelengths.)

response is characteristic of the nature of the photo-surface. The incoming light distribution may be further modified by a multiplicative function which is the transmission characteristic of an optical filter so that the effective sensitivity $T_\lambda{}^c$, where c may stand for camera tube and filter combinations C_1, C_2 or C_3, is the product of the filter transmission and camera photo-surface sensitivity. Where, as in usual ideas for either the simultaneous or sequential color camera and transmitting equipment, each camera tube and filter combination has sole control over one receiver primary, so that each C may be identified with one B, the effective sensitivities may be written as $T_\lambda{}^b$.

To satisfy the unique colorimetric requirement on the camera, therefore, it must be that $T_\lambda{}^b = I_\lambda{}^b$, excepting for constant proportionality factors which apply in effect to the amplification gains in the three channels. Herein lies the difficulty with the idea of the separate derivation of the three signals. The functions I_λ^b must necessarily be negative over some part of the wavelength spectrum. (Figure 3.) This is a logical consequence of the colorimetric relationships and is most easily seen in relation 6, where it must be remembered that the receiver primaries here considered are perfectly real light sources and have only positive spectral intensity distributions D_b^λ. Thus, for example, the requirement that $\int_{380}^{750} I_\lambda^{B_1} D_{B_2}^\lambda \, d\lambda = 0$ means, since $D_{B_2}^\lambda$ is always positive and of appreciable amplitude over a considerable range of wavelengths and since $I_\lambda^{B_1}$ is, as shown before, a linear combination of the linearly independent functions x, y and z, and hence can be zero only at a few discrete points, that $I_\lambda^{B_1}$ must have some negative values. To provide an intuitive background for this fact it should be remembered that on the chromaticity diagram, colors outside the triangle of a primary set required negative amounts of at least one of the primaries for a match, and monochromatic colors came under this category. The television camera operates along this procedure, that it finds out the amount of the receiver primaries required to match each of the monochromatic wavelengths in the sample and then adds up the result, and thus along the way these negative amounts have to be taken into account even though the final sum represents a color which is well within the triangle of the primaries.

While by the choice of suitable optical filters it is possible to approach the positive portions of the curves I_λ^b with the camera sensitivities T_λ^b, as yet no camera tubes have been developed which yield, in effect, negative photo-response in some wavelength intervals. Hence it has not been possible to realize the negative portions of the I_λ^b functions which are essential to accurate color reproduction when the three receiver signals are derived separately. The frequently suggested remedy, that a constant or bias be added to the I_λ^b functions so as to make them positive the camera sensitivities be adjusted to the result, and then subsequently a constant signal be subtracted, is of course useless because the 'constant' signal is a function of the incoming light characteristic. On the other hand, if there were provided another camera tube for each of the ones in the conventional arrangements, such that the effective sensitivity of one corresponds to the positive part of the I_λ^b, and the sensitivity of the other to the negative part, the differences of the outputs would give precisely the desired signals.

In contrast to the cumbersome, objectionable ways available to attain the camera spectral sensitivities I_λ^b necessary to accurate color reproduction when the receiver signals are each derived by a separate photo-pickup arrangement, an elegant method of deriving the signals in simultaneous systems becomes evident upon a more thoroughgoing application of the colorimetric relationship. From relation 3 it is apparent that the $I_\lambda{}^b$ may be expressed as independent linear combinations of analogous functions applying to other primary sets, as for instance the C set of primaries later to be identified with the separate camera tubes: $I_\lambda{}^b = I_c{}^b I_\lambda{}^c$. Similarly the $I_\lambda{}^c$ may be expressed as linear combinations of $I_\lambda{}^a$ where the A set of primaries are the standard ideal set and hence the $I_\lambda{}^a$ are also the conventional $\bar{x}$, $\bar{y}$ and $\bar{z}$. Now if three camera tubes with effective sensitivities $T_\lambda{}^c = I_\lambda{}^c$ are used to scan the sample color, their separate outputs will consist of the integrals $EI_\lambda{}^c D_s{}^\lambda$ just as in the conventional schemes in which the receiver contained the primary set C. These signals then are combined by algebraic addition in three different ways by linear networks in accordance with the three transformations defined by the coefficients $I_c{}^b$ which are fixed into the network by potentiometers or the like. The output signals thus have the magnitude $I_c{}^b EI_\lambda{}^c D_s{}^\lambda$ which are exactly the same as the desired signals $EI_\lambda{}^b D_s{}^\lambda$. Clearly the C set of primaries have played an entirely intermediate role and they do not have to be actually realized. There is no requirement therefore that their spectral intensity distributions $D_c{}^\lambda$ be entirely positive quantities and consequently there is no reason why the functions $I_\lambda{}^c$ have to have some negative values. The only requirement on the $I_\lambda{}^c$ is that they be independent linear combinations of the positive functions $\bar{x}$, $\bar{y}$ and $\bar{z}$. Hence there is comparative freedom to choose these functions $I_\lambda{}^c$, first so that they are positive and make possible the design of optical filters which will give the proper camera sensitivities $T_\lambda{}^c = I_\lambda{}^c$ and yet so the subtraction of signals which occurs in the combining networks represents at least for white light, a comparatively small correction on the main signal strength going through each channel. This is asked because in subtraction, signal strengths may be diminished but noise always increases. Finally, in the interests of convenience the combinations of $\bar{x}$, $\bar{y}$ and $\bar{z}$ should be chosen such that, when the spectral sensitivities of the photo-surfaces are divided out, the resulting filter characteristics which are required can be secured with obtainable dyes.

In the design of the light filters to provide the camera tubes with the effective sensitivities $T_\lambda{}^c$ which have been selected, additional characteristics are frequently incorporated to compensate for the light source illuminating the subject being televised. It very often happens that the light source is limited in its spectral characteristics by other considerations, as for example in the flying spot slide and movie scanner where the phosphor must be of high intensity and very short decay time, even though slides and movies are ordinarily intended for projection with light from a tungsten lamp. Rather than applying an optical filter directly to the light from the phosphor, a very substantial saving in intensity can be obtained by modifying the effective sensitivities of the camera tubes so as to give the equivalent overall characteristics as if the proper illuminant were used.

The network used for the algebraic addition of the camera signals usually requires some means of phase inversion to secure at the same time signals of both polarities. (See Figure 4.) The combination of the signals is then easily accomplished by purely resistive elements which may have already fixed in them the appropriate constants I_c^b as computed from colorimetric relations 2 and 4 and the camera sensitivity and receiver intensity distributions I_λ^c and D_b^λ. For reasons of avoiding

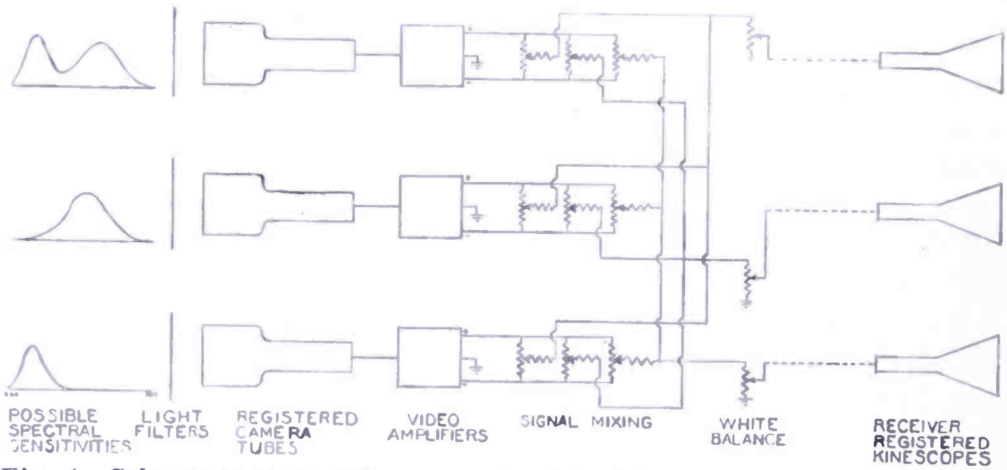


Fig. 4—Schematic of simultaneous color television system capable of exact color reproduction.

these calculations and of avoiding the need for reliance on precision in the circuit components, it is perhaps better to provide adjustable circuit elements which can be set when the system is in operation. In spite of the large number of variables, usually nine or more, a rapidly converging procedure of adjustment can be secured through the obvious facts that the transmission of white must yield white and that the camera, when viewing a light source[6] of the same tristimulus values as any one of the three primaries in the receiver, together with the combining networks, must yield a signal in the corresponding channel controlling that primary, and in no other channel. The adjustment of the white requires a balancing of the three signals output to the receiver and for this purpose it is more convenient to provide separate gain controls in the output channels even though there is then a duplication of variables.

[6] This source may conveniently be a dummy receiver.

As a result of the rigorous application of colorimetric information to simultaneous color television which is admirably well suited to the purpose, a new medium for the reproduction of color is becoming available. It is capable of by far the finest performance yet known in commercial processes, having at the same time a wide gamut of colors, colors of very high saturation, and an intrinsically accurate means of adjusting these colors automatically.

References

Visual Properties pertaining to color: W. D. Wright, RESEARCHES ON NORMAL AND DEFECTIVE COLOUR VISION, Kimpton, London, 1946.

Colorimetry: Massachusetts Institute of Technology, HANDBOOK OF COLORIMETRY, The Technology Press, Cambridge, 1936.

Color Reproduction: A. C. Hardy and F. L. Wurzburg, "Theory of Three Color Reproduction", *Jour. Opt. Soc. Amer.*, Vol. 27, p. 227, 1937.

Simultaneous Color Television: A progress report, "Simultaneous All-Electronic Color Television", *RCA REVIEW*, Vol. VII, No. 4, pp. 459-468, December, 1946.

R. D. Kell, "An Experimental Simultaneous Color Television System—Part I, Introduction"; G. C. Sziklai, R. C. Ballard and A. C. Schroeder, "Part II, Pick-up Equipment"; K. R. Wendt, G. L. Fredendall and A. C. Schroeder, "Part III, R-F and Reproducing Equipment," *Proc. I.R.E.*, Vol. 35, No. 9, September, 1947.

OPTIMUM RESISTIVE TERMINATIONS FOR SINGLE-SECTION CONSTANT-K LADDER-TYPE FILTERS*

By

L. J. Giacoletto

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—*The operation of a single non-dissipative section of a ladder-type constant-K filter terminated in a resistance is considered. It is found that depending upon the value of the terminating resistance in proportion to a filter-design parameter, different filter characteristics are obtained. Optimum values for the resistive termination are determined for different operating characteristics and different filter sections. It is found that the T-filter section has somewhat better operating characteristics than the π-filter section.*

Introduction and General Considerations for T-Low Pass Filter

Consider a T-low pass filter section as shown in Figure 1. Following the usual filter design[1] the parameters in Figure 1 are given by

$$L_1 = \frac{R}{\pi f_c} \quad (1) \qquad\qquad C_2 = \frac{1}{\pi f_c R} \quad (2)$$

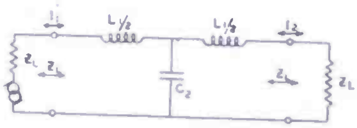


Fig. 1—T-Low pass filter section.

where f_c is the cut-off frequency and R is a design parameter equal to the input impedance of the filter at zero frequency when it is terminated by an equal impedance, or otherwise defined as the iterative impedance at zero frequency.

The input impedance of the filter section shown in Figure 1 when operated on an iterative impedance basis[2] varies with frequency in the

* Decimal Classification: R143.2 × R386.

[1] See, for example, F. E. Terman, RADIO ENGINEERS' HANDBOOK, (p. 228), McGraw-Hill Book Company, New York, N. Y., 1943.

[2] Terman, see Reference 1: "A network operating on an iterative impedance basis requires that the load impedance be such that the input impedance of the network with the load connected is equal to the load impedance." (p. 206.)

manner shown in curve 1 of Figure 2a. The same network operating on an iterative impedance basis has an attenuation as a function of frequency as given in curve 1 of Figure 2b and a phase shift as a function of frequency as given in curve 1 of Figure 2c. The iterative impedance, iterative attenuation, and iterative phase shift are given by

$$Z_I/R = \sqrt{1-u^2} \quad (3)$$

$$\alpha = \cosh^{-1} |\, 1 - 2u^2 \,| \text{ nepers } (u \geqq 1.0)\,;$$

$$= 8.686 \cosh^{-1} |\, 1 - 2u^2 \,| \text{ decibels } (u \geqq 1.0)\,; = 0 \; (u \leqq 1.0) \quad (4)$$

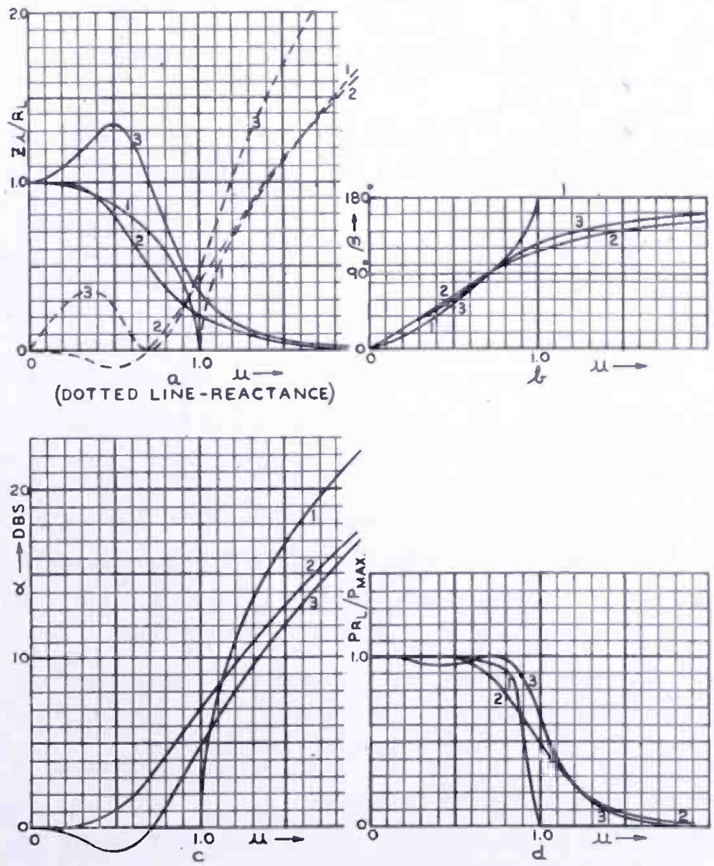


Fig. 2—Characteristics of a T-filter section.

$$\beta = \cos^{-1} (1 - 2u^2) \; (u \leqq 1.0)\,; = \pi \text{ radians } (u \geqq 1.0) \quad (5)$$

where $u = \dfrac{f}{f_o}$ (6), $\quad e^{-\theta} = e^{-(\alpha + j\beta)} = \dfrac{I_2}{I_1}$ (7)

In normal practice it is not feasible nor necessarily desirable to operate a filter section on an iterative impedance basis. In actual usage it is frequently found that the filter is required to work between a generator with a fixed resistive impedance of R_L and an equal load of R_L. Assuming that L_1 and C_2 for this filter section are again designed using (1) and (2) (including the design parameter, R), the input impedance, attenuation, and phase shift are given by

$$Z_i/R_L = \frac{1}{(2uk)^2 + (2u^2-1)^2} + j\frac{2u(2u^2-1)}{k}\left[\frac{u^2+k^2-1}{(2uk)^2+(2u^2-1)^2}\right] \tag{8}$$

$$\alpha = 1/2 \log_e [(2uk)^2 + (2u^2-1)^2] \text{ nepers}$$

$$= 10 \log_{10} [(2uk)^2 + (2u^2-1)^2] \text{ decibels} \tag{9}$$

$$\beta = -\tan^{-1}\left[\frac{2uk}{2u^2-1}\right] \tag{10}$$

where u is given in (6), α and β defined by (7) and k is a new design parameter defined as

$$k = \frac{R_L}{R} \tag{11}$$

If $k = 1.0$, that is, if the load impedance is equal to the design iterative impedance, which is the condition usually realized when following customary filter design procedure, then the input impedance, attenuation, and phase shift vary with frequency as indicated in curve 2 of Figures 2a, b, and c, respectively. If $k = 0.7$ the corresponding quantities are given by curve 3 of the same figures.

Consider for the moment the resistive component of the input impedance; it is readily seen that there is some value of k between $k = 0.7$ and $k = 1.0$ for which the generator impedance is most uniformly matched over a given frequency range, say, for example, over the pass band. That is, in order to match most uniformly the generator impedance to the resistive component of the input impedance throughout the pass band, the filter should be designed using a value of the design parameter R somewhat greater than the actual value of the load resistance, R_L. Similarly by placing other criteria such as most linear phase shift in the pass band or minimum reactance in the pass band, an optimum value for the design parameter k can again be found. Optimum values for k for different filter operating characteristics will be determined.

Power Considerations

Frequently the amount of power delivered to the filter load is of more interest than the attenuation of voltage or current. In order to determine this quantity consider a constant voltage sinusoidal generator, voltage V in series with its internal impedance, R_L, and load impedance, Z_i. Let the load impedance be normalized in terms of the generator impedance as $Z_i/R_L = R_N + j\,X_N$ (12)

Then if V is the peak voltage of the generator the average power delivered to the load is

$$P_{R_i} = 1/2 \frac{V^2 R_N}{R_L\,[(1+R_N)^2 + X_N{}^2]} \tag{13}$$

The maximum power that can be delivered to R_i is when $R_i = R_L$ ($R_N = 1.0$) and $X_N = 0$ under which condition the maximum power is

$$P_{\max} = \frac{V^2}{8\,R_L} \tag{14}$$

If Z_i is the input of a filter section any power delivered to the input of the filter must in turn be delivered to the terminating resistance since the elements are assumed to be ideal pure reactances. Therefore $P_{R_i} = P_{R_L}$ and (13) can be written

$$\frac{P_{R_L}}{P_{\max}} = 4 \frac{R_N}{[(1+R_N)^2 + X_N{}^2]} \tag{15}$$

For a T-low pass filter operated on an iterative impedance basis the power relation to accompany Equations (3), (4), and (5) is

$$\frac{P_{R_L}}{P_{\max}} = 4 \frac{\sqrt{1-u^2}}{[1+\sqrt{1-u^2}]^2}\ (u \leqq 1.0)\,;\ = 0\ (u \geqq 1.0) \tag{16}$$

This curve is shown as curve 1 in Figure 2d. For the filter operating into a terminating impedance R_L the power relation to accompany Equations (8), (9), and (10) is

$$\frac{P_{R_L}}{P_{\max}} = 4 \frac{k^2\,[(2uk)^2 + (2u^2-1)^2]}{k^2\,[1 + (2uk)^2 + (2u^2-1)^2]^2 + [2u\,(2u^2-1)\,(u^2+k^2-1)]^2} \tag{17}$$

For $k = 1.0$ this relation is shown in curve 2 of Figure 2d and in curve 3 of the same figure for $k = 0.7$.

T-Filter Uniform Resistive Impedance Throughout Pass Band

It is desired to have the input resistance of the T-filter be equal to R_L as nearly as possible throughout the pass band. The standard error* will be used as an index of the uniformity of the input resistance. Therefore,

$(\text{Standard Error}_R)^2 =$

$$(\text{S.E.}_R)^2 = \int_0^1 \left[1 - \frac{1}{(2uk)^2 + (2u^2 - 1)^2}\right]^2 du \qquad (18)$$

The optimum value of k is determined by finding the value of k for which the standard error (or what amounts to the same thing, the square of the standard error) is a minimum. Mathematically this procedure requires the evaluation of the integral in (18) and the determination of the value of k for which the derivative of this integral with respect to k is zero. The direct mathematical solution becomes prohibitively complex,† so that the optimum value of k was determined by evaluating the integral of (18) graphically for several values of k.

* The standard error ($S.E.$) is defined statistically as the square root of the average of the summation of the deviations squared. In this case, since the deviations are distributed continuously the integral is used for the summation and the interval $\int_0^1 du$ is unity.

† By means of partial fractions $(S.E._R)^2$ can be evaluated as

$$(\text{S.E.}_R)^2 = 1 + \frac{1}{4(A-B)}\left[\frac{1}{\sqrt{A}}\log_e\left(\frac{\sqrt{A}+1}{\sqrt{A}-1}\right) - \frac{1}{\sqrt{B}}\log_e\left(\frac{\sqrt{B}+1}{\sqrt{B}-1}\right)\right]$$

$$+ \frac{1}{32(A-B)^3}\left[\frac{A-B}{A(A-1)} - \frac{B-A}{B(B-1)} + \frac{5A-B}{2A\sqrt{A}}\log_e\left(\frac{\sqrt{A}+1}{\sqrt{A}-1}\right) - \frac{5B-A}{2B\sqrt{B}}\log_e\left(\frac{\sqrt{B}+1}{\sqrt{B}-1}\right)\right]$$

where

$A = -\tfrac{1}{2}\left[(k^2-1) + \sqrt{(k^2-1)^2 - 1}\right]$ and

$B = -\tfrac{1}{2}\left[(k^2-1) - \sqrt{(k^2-1)^2 - 1}\right]$.

When $k = 0.75$ this expression gives $S.E._R = 0.269$.

When this is done and the results plotted, a curve of the type shown in Figure 3a results. From this curve it is seen that the standard error is a minimum approximately when $k = 0.75$. The variation of the input resistance for the optimum value of $k = 0.75$ is shown in Figure 3b.

Thus, it is seen that for the most uniform input impedance throughout the pass band for a T-filter the optimum design parameter is

$$k = \frac{R_L}{R} = 0.75 \tag{19}$$

and in order to design a filter operating under these conditions it is necessary to employ a design parameter R which is 4/3 as great as the load resistance used to terminate the T-filter. That is,

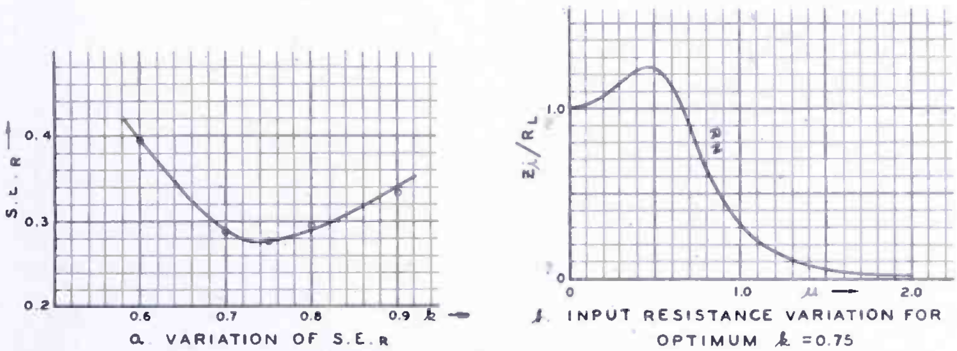


Fig. 3—T-Filter uniform resistive impedance.

$$R = \frac{4}{3} R_L \tag{20}$$

When this is done, the variation with frequency of the other characteristics of the filter can be ascertained approximately by examination of Figure 2 or exactly by using $k = 0.75$ in Equations (8), (9), (10), and (17).

T-Filter Minimum Reactive Impedance Throughout Pass Band

In this case it is desired that the input reactive impedance of the T-filter be minimum throughout the pass band. The standard error is

$$(S.E._X)^2 = \int_0^1 \left\{ \frac{2u\,(2u^2 - 1)}{k} \left[\frac{u^2 + k^2 - 1}{(2uk)^2 + (2u^2 - 1)^2} \right] \right\}^2 du \tag{21}$$

The evaluation of the integral is again done graphically with the results shown in Figure 4a. The standard error has a minimum with the optimum k being approximately

$$k = \frac{R_L}{R} = 0.95 \tag{22}$$

The variation of the input reactance for the optimum value of $k = 0.95$ is shown in Figure 4b.

T-Filter Linear Phase Shift Throughout Pass Band

In this case it is desired that the phase shift vary linearly[3] throughout the pass band. The standard error is

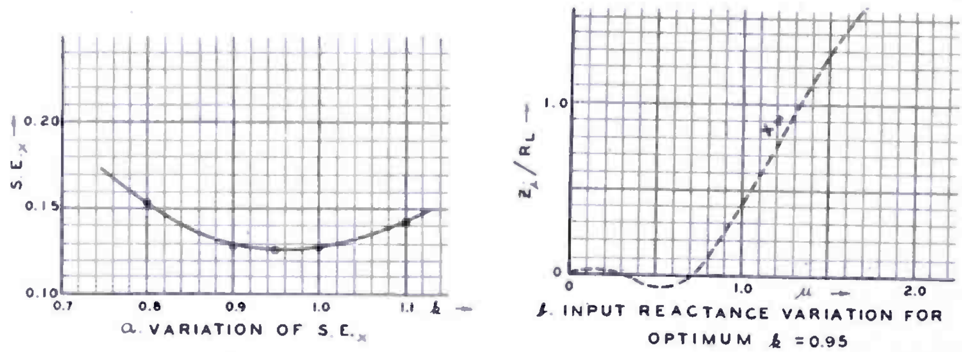


Fig. 4—T-Filter minimum reactive impedance.

$$(S.E._{\beta})^2 = \int_0^1 \left\{ Mu - \tan^{-1} \left[\frac{2uk}{2u^2 - 1} \right] \right\}^2 du \tag{23}$$

where M is the slope of the line that is to be used as the standard from which deviations are computed. The problem is therefore complicated by the fact that M can be given any value desired for different values of k. The procedure followed was to plot β curves for several different values of k and through each of these curves a straight line that would appear by visual inspection to give the least standard deviation is drawn. It is then found that in the vicinity of the optimum k the straight line that comes the closest to fitting the β curves all have the

[3] In video amplifiers it is necessary to have a constant time delay for all frequencies to be amplified in order to prevent distortion. This necessitates a linear phase shift with frequency. Filter networks have been used both as coupling networks and as plate loads in video amplifiers. See H. A. Wheeler, "Wide Band Amplifiers for Television," *Proc. I.R.E.*, Vol. 27, No. 7, pp. 429-437; July, 1939. In some applications constant time delay throughout the pass band may be a more important criterion of operation; in this case a different optimum k value is to be expected.

same value of M, namely 123.0 degrees per unit value of u. Using the indicated value of M, several values of the standard error were computed graphically and the results plotted in Figure 5a. This figure indicates a minimum in the vicinity of

$$k = \frac{R_L}{R} = 0.97 \tag{24}$$

The variation of phase shift for the optimum value of $k = 0.97$ is shown in Figure 5b where the dotted line is the assumed standard of slope M. It is worth noting that all β curves pass through the value $\beta = 90$ degrees when $u = \dfrac{1}{\sqrt{2}}$. A small distance beyond this value the

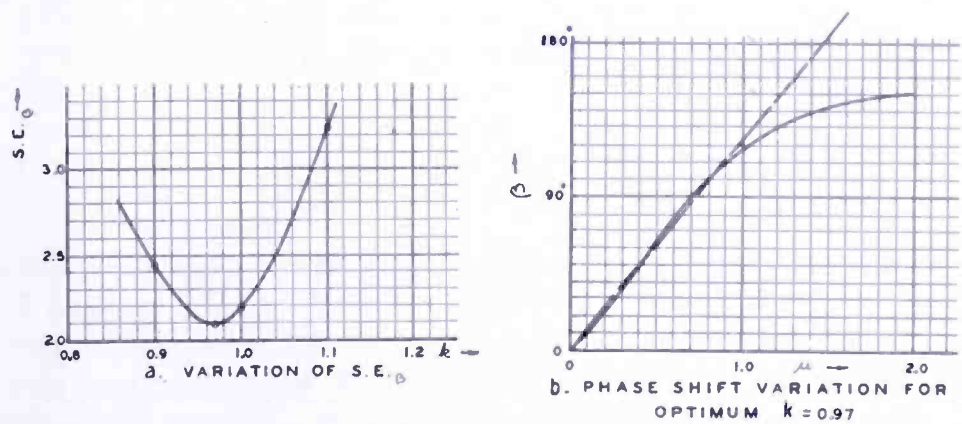


Fig. 5—T-Filter linear phase shift.

phase shift curves began to depart rather widely from the straight line variation.

T-Filter Maximum Power Throughout Pass Band

It is sometimes desirable to have the filter deliver the maximum amount of power to the terminating load when operating throughout the pass band. For this case, the standard error using Equation (17) is

$$(S.E._P)^2 = \int_0^1 \left\{ 1 - 4 \frac{k^2 \left[(2uk)^2 + (2u^2 - 1)^2 \right]}{k^2 \left[1 + (2uk)^2 + (2u^2-1)^2 \right]^2 + \left[2u\,(2u^2-1)\,(u^2+k^2-1) \right]^2} \right\}^2 du \tag{25}$$

This integral was evaluated graphically with the results shown in

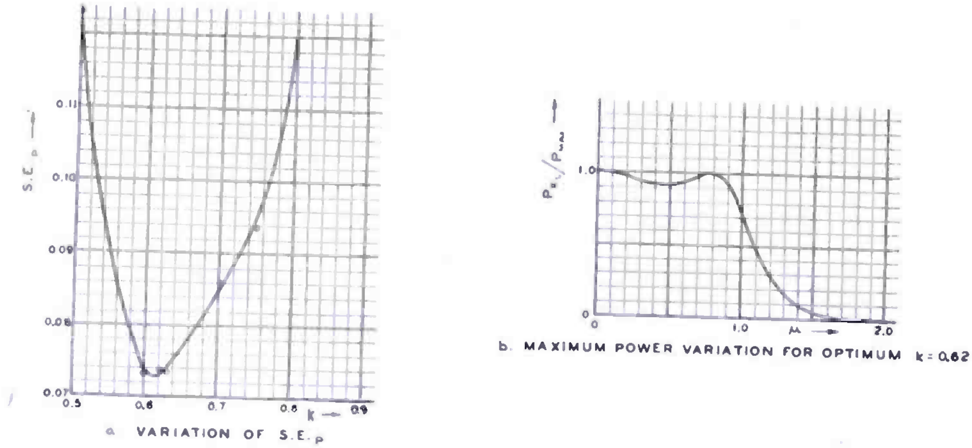


Fig. 6—T-Filter maximum power.

Figure 6a. This figure indicates a minimum in the vicinity of

$$k = \frac{R_L}{R} = 0.62 \tag{26}$$

The variation of the load power to the maximum power for the optimum value of $k = 0.62$ is shown in Figure 6b. The optimum value of k is not the value that most uniformly meets the desired condition over a limited frequency range. Thus, for $k = 0.75$ (the same value as for uniform input resistance), the value of P_{R_L}/P_{max} is very nearly unity from $u = 0$ to $u = 0.8$. However, from $u = 0.8$ to $u = 1.0$ the curve slumps to such an extent that the standard error in this case is greater than for the optimum value of $k = 0.62$. However, for some usages the former value of $k = 0.75$ may be more desirable particularly since this value also gives the most uniform variation of input resistance through the pass band.

General Considerations for π-Low Pass Filter

For the π-low pass filter indicated in Figure 7 and following the usual filter design, the parameters are again given by Equations (1) and (2). The input impedance, attenuation, phase shift and power delivered to the load when operated on an iterative impedance basis

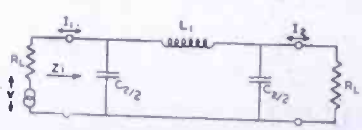


Fig. 7—π-Low pass filter section.

vary with frequency in the manner shown in curve 1 of Figures 8a, b, c, and d, respectively. The analytical expression for these quantities comparable to Equations (3), (4), (5), and (16) for the T-low pass filter are

$$Z_i/R = \frac{1}{\sqrt{1-u^2}} \tag{27}$$

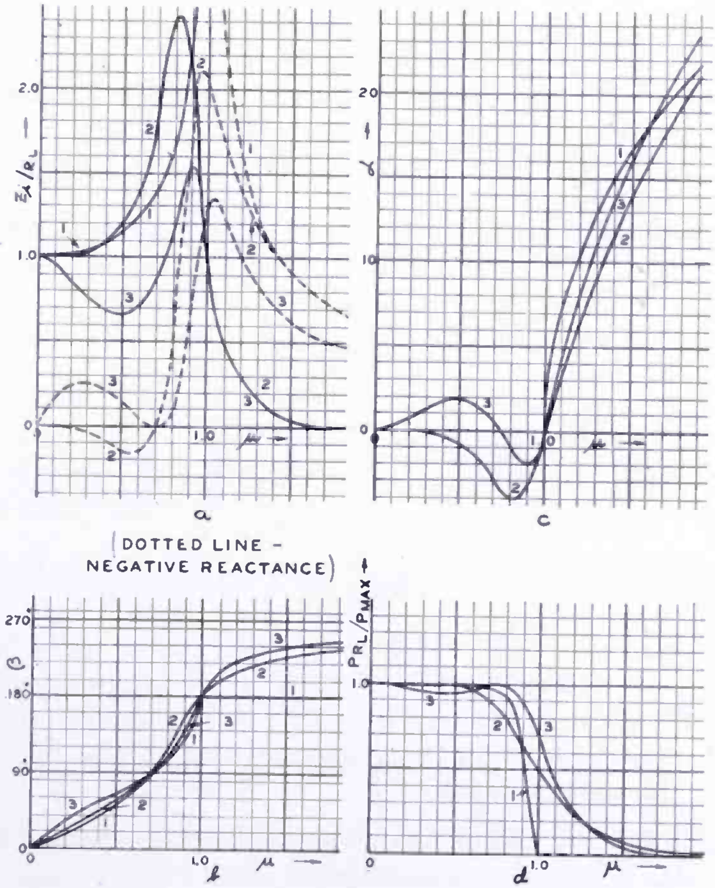


Fig. 8—Characteristics of a π-filter section.

$$\alpha = \cosh^{-1}|1-2u^2| \text{ nepers } (u \geqq 1.0); \tag{28}$$

$$= 8.686 \cosh^{-1}|1-2u^2| \text{ decibels } (u \geqq 1.0); = 0 \ (u \leqq 0).$$

$$\beta = \cos\ (1-2u^2)\ (u \leqq 1.0); = \pi \text{ radians } (u \geqq 1.0) \tag{29}$$

$$P_{R_L}/P_{\max} = 4\frac{\sqrt{1-u^2}}{[1+\sqrt{1-u^2}]^2}\ (u \leqq 1.0); = 0\ (u \geqq 1.0). \tag{30}$$

where u has the value given in (6) and α and β are defined by (7).

With L_1 and C_2 computed by the use of (1) and (2), the input impedance, attenuation, phase shift, and power delivered to the load are given by the analytic expressions

$$Z_i/R_L = \left[\frac{1}{(2uk)^2(u^2-1)^2+(2u^2-1)^2}\right]$$

$$-j\frac{2u(2u^2-1)}{k}\left[\frac{k^2(u^2-1)+1}{(2uk)^2(u^2-1)^2+(2u^2-1)^2}\right] \quad (31)$$

$$\alpha = 1/2\log_e[(2uk)^2(u^2-1)^2+(2u^2-1)^2] \text{ nepers}$$

$$= 10\log_{10}[(2uk)^2(u^2-1)^2+(2u^2-1)^2] \text{ decibels} \quad (32)$$

$$\beta = \tan^{-1}\left[\frac{2uk(u^2-1)}{2u^2-1}\right] \quad (33)$$

$$P_{R_L}/P_{max} = 4\frac{(2uk)^2(u^2-1)^2+(2u^2-1)^2}{[\{1+(2uk)^2(u^2-1)^2+(2u^2-1)^2\}^2 + \frac{4u^2}{k^2}(2u^2-1)^2\{k^2(u^2-1)+1\}^2]} \quad (34)$$

which correspond to Equations (8), (9), (10), and (17), respectively, for the T-low pass filter. In the above equations k has the same value defined in (11), namely R_L/R.

If $k = 1.0$ then the quantities given above are indicated by curve 2 of Figures 8a, b, c, and d, respectively. For $k = 1.5$ these quantities are indicated by curve 3 of the same figures.

π-Filter Uniform Resistive Impedance Throughout Pass Band

It is desired to have the input resistance of the π filter be equal to R_L as nearly as possible throughout the pass band. The standard error in this case is

$$(\text{Standard Error}_R)^2 = (S.E._R)^2 =$$

$$\int_0^1\left[1-\frac{1}{(2u^2-1)^2+(2uk)^2(u^2-1)^2}\right]^2 du \quad (35)$$

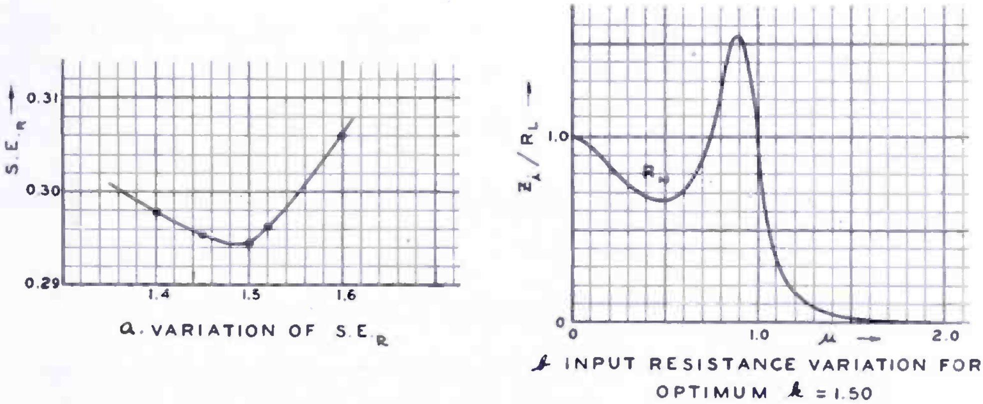


Fig. 9—π-Filter uniform resistive impedance.

The results of the evaluation of this integral by graphical methods is shown in Figure 9a and indicates that the standard error has a minimum when

$$k = \frac{R_L}{R} = 1.50 \tag{36}$$

The variation of the input resistance for this optimum value of k is shown in Figure 9b.

π-Filter Minimum Reactive Impedance Throughout Pass Band

It is desired to have the input reactive impedance of the π filter be a minimum throughout the pass band. The standard error in this case is

$$(S.E._X)^2 = \int_0^1 \left[-\frac{2u\,(2u^2-1)\,[k^2\,(u^2-1)+1]}{k\,[(2uk)^2\,(u^2-1)^2+(2u^2-1)^2]} \right]^2 du \tag{37}$$

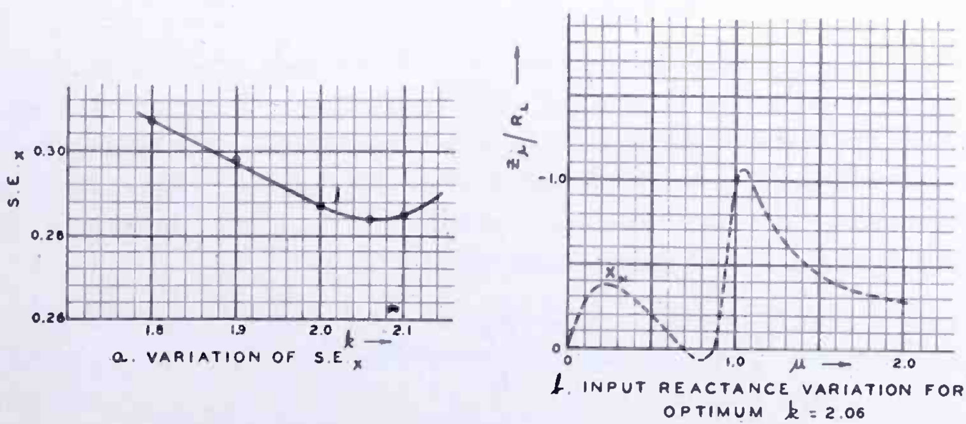


Fig. 10—π-Filter minimum reactive impedance.

Figure 10a indicates the results of the graphical integration with a minimum when

$$k = \frac{R_L}{R} = 2.06 \tag{38}$$

The variation of input reactance for this optimum value of k is shown in Figure 10b.

π-Filter Linear Phase Shift Throughout Pass Band

It is desired to have the phase shift vary as nearly linearly as possible throughout the pass band. The standard error is

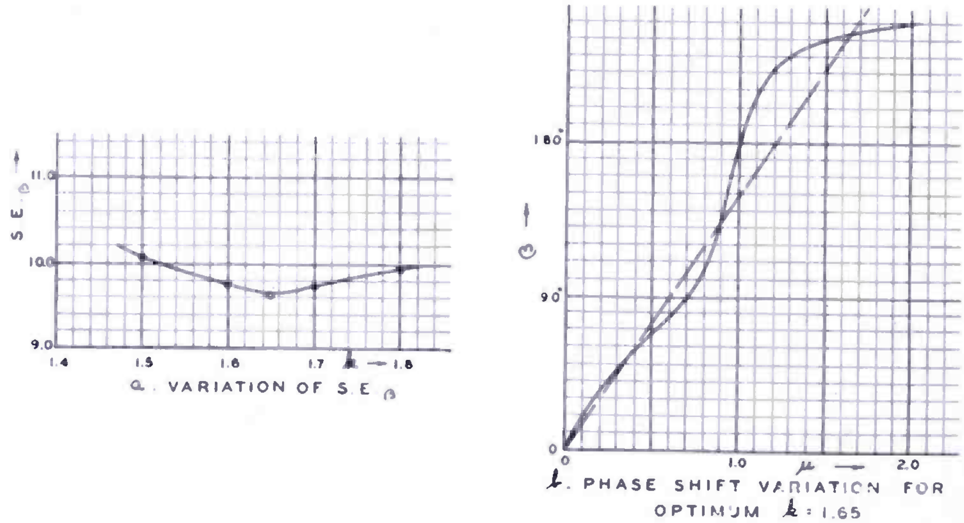


Fig. 11—π-Filter linear phase shift.

$$(S.E._\beta)^2 = \int_0^1 \left\{ \mathrm{M}u - \tan^{-1}\left[\frac{2uk\ (u^2-1)}{2u^2-1}\right]\right\}^2 du \tag{39}$$

where M is the slope of the line that is to be used as a standard from which deviations are computed. The procedure followed was the same as for the T filter except in this case the best value for M was found to be 148 degrees per unit value of u. Using this value of M, the standard error when evaluated graphically is given in Figure 11a. This figure indicates a minimum when

$$k = \frac{R_L}{R} = 1.65 \tag{40}$$

The variation of β for this optimum value of k is shown in Figure 11b.

All β curves pass through the value $\beta = 90$ degrees when $u = \dfrac{1}{\sqrt{2}}$ and through $\beta = 180$ degrees when $u = 1.0$.

π-Filter Maximum Power Throughout Pass Band

The standard error in this case is

$$(S.E._P)^2 = \int_0^1 \left[1 - 4 \frac{(2uk)^2\,(u^2-1)^2 + (2u^2-1)^2}{[\{1 + (2uk)^2\,(u^2-1)^2 + (2u^2-1)^2\}^2} \right.$$

$$\overline{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad} \qquad (41)$$

$$\left. + \frac{4u^2}{k^2}\,(2u^2-1)^2\,\{k^2\,(u^2-1)+1\}^2] \right]^2 du$$

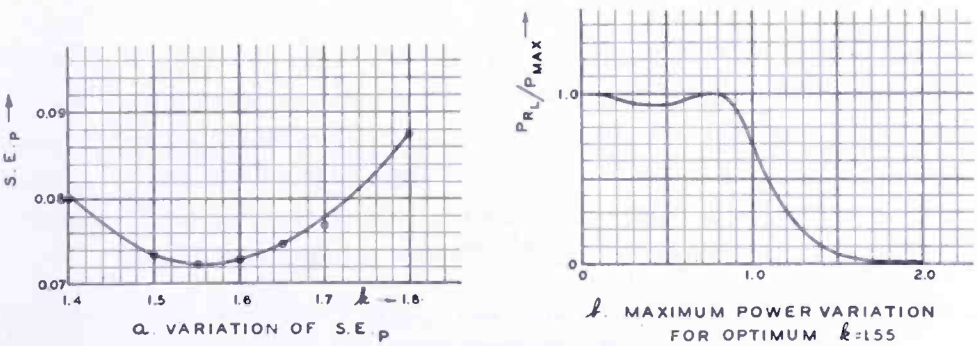


Fig. 12—π-Filter maximum power.

The results of graphical evaluation of this standard error are shown in Figure 12a which indicates that a minimum exists in the neighborhood of

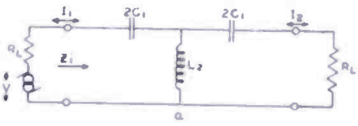
$$k = \frac{R_L}{R} = 1.55 \qquad (42)$$

The variation of the load power to the maximum power for this optimum value is shown in Figure 12b.

T- and π-High Pass Filter

For the T- and π-high pass filter sections shown in Figures 13a and b, respectively, following the usual filter design the parameters are given by

$$C_1 = \frac{1}{4\pi f_c R} \qquad (43) \qquad\qquad L_2 = \frac{R}{4\pi f_c} \qquad (44)$$



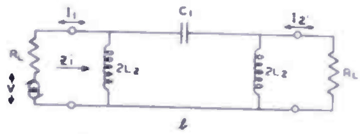


Fig. 13—High pass filter sections.

where f_c is the cut-off frequency and R is the same design parameter as for the low pass filter sections. Using (43) and (44) the equations for the input impedance, attenuation, phase shift, and power to the load for T and π high pass filter sections operated on an iterative impedance basis or terminated in a load resistance R_L come out to be the same as corresponding formulas for T- and π-low pass filters provided u is replaced by v defined as

$$v = \frac{f_c}{f} \tag{45}$$

and the sign of the reactance and phase shift is changed. The latter changes require that $+j$ be replaced by $-j$ and $+\beta$ be replaced by $-\beta$. Thus, it is seen that all optimum values of k determined for T- and π-low pass filter are valid for T- and π-high pass filters and all curves plotted of T- and π-low pass filter characteristics can be easily changed to corresponding characteristics for T- and π-high pass filters.

T-Band Pass Filter

For the T-band pass filter indicated in Figure 14 the parameters

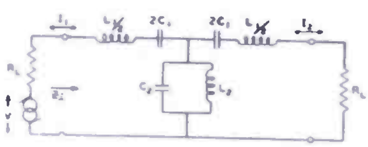


Fig. 14—T-Band pass filter section.

are given as[4]

$$L_1 = \frac{R}{\pi\,(f_2 - f_1)} \tag{46}$$

$$C_1 = \frac{f_2 - f_1}{4\pi\, f_1\, f_2\, R} \tag{47}$$

$$L_2 = R^2 C_1 = \frac{(f_2 - f_1)\, R}{4\pi\, f_1\, f_2} \tag{48}$$

$$C_2 = \frac{L_1}{R_2} = \frac{1}{\pi\,(f_2 - f_1)\, R} \tag{49}$$

[4] See Knox McIlwain and J. G. Brainerd, HIGH FREQUENCY ALTERNATING CURRENTS, p. 303, John Wiley and Sons, Inc., New York, N. Y., 1931.

where f_1 is the lower cut-off frequency of the pass band and f_2 is the upper cut-off frequency of the pass band. If Z_1 is the series impedance of L_1 in series with C_1 and Z_2 the shunt impedance of L_2 in parallel with C_2, then

$$Z_1 = j2R \frac{\sqrt{f_1 f_2}}{(f_2 - f_1)} \left[\frac{f}{\sqrt{f_1 f_2}} - \frac{\sqrt{f_1 f_2}}{f} \right] = j2Rw \tag{50}$$

$$Z_2 = -j \frac{R\,(f_2 - f_1)}{2\sqrt{f_1 f_2}} \left(\frac{1}{\dfrac{f}{\sqrt{f_1 f_2}} - \dfrac{\sqrt{f_1 f_2}}{f}} \right) = -j \frac{R}{2w} \tag{51}$$

where w is a new frequency parameter defined as

$$w = \frac{f_m}{f_2 - f_1} \left[\frac{f}{f_m} - \frac{f_m}{f} \right] \tag{52}$$

and f_m is the geometric mean frequency between the two cut-off frequencies

$$f_m = \sqrt{f_1 f_2} \tag{53}$$

and also the frequency at which the series impedance is zero and the shunt impedance is infinite.

By use of the new frequency parameter w, the T-band pass filter becomes identical in operating characteristics with a T-low or high pass filter so that the equations governing the operation of the T-band pass filter on an iterative impedance basis are given by Equations (3), (4), (5), and (16) with u replaced by w and the operation of the T-band pass filter when terminated in a fixed resistance R_L is defined by Equations (8), (9), (10), and (17) with u again replaced by w and k defined by (11).

It therefore follows that the optimum design value of k for the different operating characteristics will be the same for the T-band pass filter as for the T-low or high pass filter. The value of the total error in each case will be twice as large for the T-band pass filter as for the T-low or high pass filter as the pass band occurs when $-1 \leq w \leq 1$. The standard error, however, will be the same since the interval $\int_{-1}^{1} dw$ is also twice as large.

π-BAND PASS FILTER

For the π-band pass filter indicated in Figure 15 the parameters are defined by (46), (47), (48), and (49). The equations for the π band pass filter operated on an iterative impedance basis are given by Equations (27), (28), (29), and (30) if u is replaced by w as defined in (52) and the equations for the π band pass filter when terminated in a fixed resistance R_L are given by Equations (31), (32), (33), and (34) with u again replaced by w and k defined by (11). Thus the optimum design value of k for the different operating characteristics as well as the standard error in each case will be the same for the π band pass filter as for the π-low or high pass filter.

T-BAND ELIMINATION FILTER

For the T-band elimination filter indicated in Figure 16 the

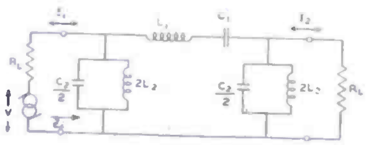


Fig. 15—π-Band pass filter section.

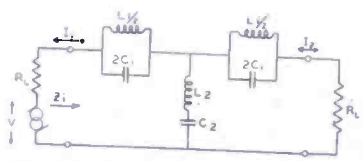


Fig. 16—T-Band elimination filter section.

parameters are given as

$$L_1 = \frac{(f_2 - f_1)\,R}{\pi\, f_1\, f_2} \qquad (54) \qquad\qquad C_1 = \frac{1}{4\pi\, R\,(f_2 - f_1)} \qquad (55)$$

$$L_2 = R^2 C_1 = \frac{R}{4\pi\,(f_2 - f_1)} \qquad (56) \qquad\qquad C_2 = \frac{L_1}{R^2} = \frac{(f_2 - f_1)}{\pi\, f_1\, f_2\, R} \qquad (57)$$

where f_1 is the lower frequency of the elimination band and f_2 is the upper frequency of the elimination band. If Z_1 is the series impedance of L_1 in parallel with C_1 and Z_2 the shunt impedance of L_2 in series with C_2, then

$$Z_1 = -j2R\,\frac{(f_2 - f_1)}{\sqrt{f_1 f_2}} \left[\frac{1}{\dfrac{f}{\sqrt{f_1 f_2}} - \dfrac{\sqrt{f_1 f_2}}{f}}\right] = -j\,\frac{2R}{w} = -j2Rx \qquad (58)$$

$$Z_2 = j\,\frac{R}{2}\,\frac{\sqrt{f_1 f_2}}{(f_2 - f_1)} \left[\frac{f}{\sqrt{f_1 f_2}} - \frac{\sqrt{f_1 f_2}}{f}\right] = j\,\frac{R}{2}\,w = j\,\frac{R}{2x} \qquad (59)$$

where x is a new frequency parameter defined as

$$x = \frac{1}{w} = \frac{1}{\dfrac{f_m}{(f_2 - f_1)} \left[\dfrac{f}{f_m} - \dfrac{f_m}{f} \right]} \tag{60}$$

where f_m is defined by (53). The pass band occurs when $-1 \leq x \leq 1$. The discussion for the T-band pass filter is applicable to the T-band elimination filter provided the change on the frequency axis is kept in mind, and provided that optimum operation is desired over the pass band rather than over the elimination band.

π-Band Elimination Filter

For the π-band elimination filter indicated in Figure 17 the

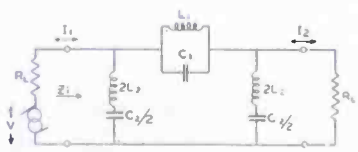


Fig. 17—T-Band elimination filter section.

parameters are defined by (54), (55), (56), and (57). The discussion for the π-band pass filter is applicable to the π-band elimination filter provided the change in the frequency axis as indicated in (60) is kept in mind, and optimum operation is again desired over the pass band.

Conclusions

The determination of optimum terminating resistances for certain operating characteristics for a single section of a constant-K ladder-type filter will enable a design engineer to obtain somewhat improved performance without the expense of additional circuit elements. Thus, the filter section is designed using conventional design formulas (see Table I) with the iterative impedance, R, being determined from the relationship $R = \dfrac{R_L}{k}$ where R_L is the desired terminating impedance (equal to the generator impedance), and the optimum values of design parameter, k, are conveniently tabulated in Table I together with the values of the standard errors. It is of interest to note that the T-filter sections give values of standard errors that are either equal to or less

Table I

TABULATION OF OPTIMUM DESIGN PARAMETER k TOGETHER WITH STANDARD ERRORS FOR THE VARIOUS CONSTANT K TYPE FILTER SECTIONS

FILTER SECTION	FIG.	DESIGN EQUATIONS	UNIFORM RESISTIVE IMPEDANCE		MINIMUM REACTIVE IMPEDANCE		LINEAR PHASE SHIFT		MAXIMUM POWER	
			$k=R_L/R$	$S.E._R$	$k=R_L/R$	$S.E._X$	$k=R_L/R$	$S.E._\beta$	$k=R_L/R$	$S.E._P$
T-Low Pass	1	1, 2	0.75	0.275	0.95	0.127	0.97	2.10	0.62	0.073
T-High Pass	13a	43, 44	"	"	"	"	"	"	"	"
π-Low Pass	7	1, 2	1.50	0.294	2.06	0.282	1.65	9.65	1.55	0.073
π-High Pass	13b	43, 44	"	"	"	"	"	"	"	"
T-Band Pass	14	46, 47, 48, 49	0.75	0.275	0.95	0.127	0.97	2.10	0.62	0.073
T-Band Elimination	15	54, 55, 56, 57	"	"	"	"	"	"	"	"
π-Band Pass	16	46, 47, 48, 49	1.50	0.294	2.06	0.282	1.65	9.65	1.55	0.073
π-Band Elimination	17	54, 55, 56, 57	"	"	"	"	"	"	"	"

than the comparable standard errors for the π-filter sections. Other things being equal then, the T-filter section is to be preferred over the π-filter section.

The method of attack used in this paper, namely that of minimizing the standard error over the pass band could be used in a similar manner for more complicated filters such as m-derived filter sections and filter sections operated in tandem. It is doubtful whether the improvement to be realized in these cases would be worth the effort as the characteristics are more nearly ideal to start with for the more complicated filters than for the single filter sections studied here.

While the preceding analysis has resulted in some useful filter design data, certain factors have not been considered. These are:

1. Filter elements have been assumed ideal whereas in actual practice some resistance is present particularly in the inductances.

2. Generator and load resistance are assumed to be pure resistances independent of frequency. In actual operation this condition is usually approximated closely although some reactance and resistance variation may be present.

3. Optimized conditions are specified for only one filter characteristic in each case without regard to what may happen to the other characteristics. In actual operation several characteristics may be of simultaneous interest. Thus, in the use of filter sections in video amplifiers not only is a linear change of phase shift with frequency desired, but also a uniform input resistance, as well as maximum power throughout the pass band. The mathematical problem of optimizing several conditions simultaneously becomes very complex mathematically. About all that can be done in this case is to make a judicious

choice between the optimum values of k for the several characteristics.

4. Filter characteristics have been optimized throughout the pass band without consideration as to what occurs outside the pass band. It is possible to make conditions outside the pass band worse by optimizing within the pass band. Thus, the condition for maximum power throughout the pass band is achieved at the expense of additional power outside of the pass band. A mathematical approach to the solution of this problem is similar to that in Section 3 above and is also complex. Fortunately, the changes in the filter characteristics outside of the pass band are small compared to the changes within the pass band so that optimizing within the pass band is permissible.

In view of the graphical method employed in evaluating the various integrals, the accuracy of the evaluation of the optimum values of k is believed to be no better than ± 2 per cent. Higher accuracy could be obtained if desired by evaluating additional points and by utilizing a more accurate method of graphical integration.

BEAM-DEFLECTION CONTROL FOR AMPLIFIER TUBES*

By

G. R. Kilgore

formerly with
Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—It is the purpose of this paper: (1) to discuss the basic principles involved in obtaining high transconductance and high ratios of transconductance to current by means of deflection control; (2) to derive expressions for the ultimate transconductance at both low and high frequencies; (3) to discuss by means of elementary electron optics the design of a simple, beam-deflection gun for obtaining the desired results; and (4) to describe some of the early experimental results on amplifier tubes combining beam-deflection control and a multi-stage secondary-emission multiplier.

It is shown that deflection control offers a possibility of obtaining substantial transconductance with low capacitance and low beam currents and with a very high ratio of transconductance to plate current. It is found experimentally that useful values of transconductance with low capacitance and low current can be obtained with a simple deflection gun combining focusing and deflection. This type of control is ideally suited for use with a high gain secondary emission multiplier to obtain very high transconductance, without excessive capacitance, thus making possible a tube with a bandwidth figure of merit many times greater than for conventional tubes.

Experimental confirmation of some of the properties of deflection control in agreement with the analysis has already been obtained in experimental tubes.

Introduction

After the successful demonstration of the secondary-emission multiplier in 1935[1] for amplifying photoelectric currents, it was natural that attempts should be made to obtain a high-transconductance, voltage-controlled amplifier with low input current. Efforts to produce such a tube with conventional grid control on the input soon indicated a basic limitation set by the ratio of transconductance to plate current. The maximum possible value of this ratio for grid control has been shown[2] to be e/kT. Thus, for ordinary

* Decimal Classification: R333 × R139.

[1] V. K. Zworykin, G. A. Morton and L. Malter, "The Secondary Emission Multiplier—A New Electronic Device," *Proc. I.R.E.*, Vol. 24, pp. 351-376, March, 1936.

[2] F. Below, "Theory of Space-Charge-Grid Tubes," *Zeit. fur Fernmeldtechnik*, Vol. 9, pp. 113-118 and 136-142, 1928.

cathodes ($T \cong 1000$ degrees Kelvin) the maximum theoretical ratio cannot exceed about 10 to 12 micromhos per microampere and, in practice, is seldom more than one or two micromhos per microampere. This meant that high transconductance could be obtained only with high current on the output stage of the multiplier and this current limit was set by dissipation and space charge, as in a conventional tube. It was clear that a new means of control was needed to give higher ratios of transconductance to plate current.[3] Deflection control appeared as one of the possibilities and a decision to explore this method was somewhat influenced by a privately communicated report on the work of Nagashima.[4] He described a complex deflection tube in which he obtained a transconductance to current ratio of 100. Although his beam current was on the order of 10^{-8} amperes and the transconductance was only 0.5 micromhos, his experiments did indicate that deflection control offered a solution to the voltage-controlled multiplier for high transconductance. Other laboratories also appreciated the advantages of deflection control for multiplier tubes[5,6,7] but the absence of further publication makes the ultimate results obtained somewhat in doubt.

Well before the war, and at the suggestion of B. J. Thompson, an investigation was begun of beam-deflection control in combination with secondary-emission multiplication with the objective of obtaining a useful device for amplification and detection of ultra-high frequencies. A particular objective of the early work was to produce a high-transconductance tube with low capacitance for wide-band amplification at ultra-high frequencies. Work on deflection control, subsequent to that covered in this paper, was continued throughout the war in these laboratories and in other divisions of RCA. Some of this later work was sponsored by the Armed Services.

The present paper covers only the early work, namely: (1) The basic principles involved in obtaining high transconductance and high ratios of transconductance to current by means of deflection control;

---

[3] B. J. Thompson, "Voltage-Controlled Electron Multipliers," *Proc. I.R.E.*, Vol. 29, pp. 583-587, November, 1941.

[4] M. Nagashima, "Voltage Multipliers," a paper presented at combined Engineering Convention at Tokyo, April, 1938.

[5] W. Flechsig and M. Sandhagen, "Electronic Amplifier Valves with Secondary Emission Multiplication," *Fernseh Hausmitt*, Vol. 2, pp. 16-25, May, 1940.

[6] F. M. Colebrook, "Ultra-Short and Decimeter-Wave Valves—Deflection of a Focused Beam as a Possible Basis for Construction," *Wireless Eng.*, Vol. 15, pp. 198-201, April, 1938.

[7] J. H. O. Harries, "Discussion of Colebrook Paper," *Wireless Eng.*, Vol. 15, pp. 324-325, June, 1938.

(2) expressions for the ultimate transconductance at both low and high frequencies; (3) elementary electron optics and the design of a simple beam-deflection gun for obtaining the desired results; and (4) some of the early experimental results on amplifier tubes combining beam-deflection control and a multi-stage secondary-emission multiplier.

Basic Principles

Beam-deflection devices for use as detectors, amplifiers, and oscillators have been described in the literature[8] by many workers dating back at least to 1906, but none of these apparently had any remarkable success. Most of the earlier workers failed because they attempted to use high-current beams with consequent space-charge difficulties and because they apparently did not appreciate the basic current-density limitation that determines transconductance.

At the beginning of the present work it was decided to use very small currents in the deflection gun and to build up the current by secondary-emission multiplication. It was recognized that high current density and not high beam current was essential in obtaining high transconductance. A very important concept for electron beams was worked out by D. B. Langmuir,[9] who pointed out that the ultimate current density that could be brought to a focus was limited by thermal velocity of the electrons and was determined by the beam voltage, the cathode current density and temperature, and by the angle of convergence of the electron beam. While in his published work Langmuir treated only the point focus case, he had extended his work to the line focus case and derived an expression very similar to that later published by J. R. Pierce.[10] Combining this maximum current density with maximum deflection sensitivity, an expression for maximum low-frequency transconductance was found which gave a basis of comparison for the later work in evolving a high transconductance gun for high frequencies. This maximum low-frequency transconductance is worked out in the next section and extended to high frequencies in the section following that.

The concept of the current density limit is so basic to the beam-deflection work that it is worth while to review the concepts in terms

[8] The earliest work is usually credited to R. Von Lieben, German Patent 179,807 (1906), and M. Dieckmann and G. Glage, German Patent 184,710 (1906).

[9] D. B. Langmuir, "Theoretical Limitations of Cathode-Ray Tubes," *Proc. I.R.E.*, Vol. 25, pp. 977-992, August, 1937.

[10] J. R. Pierce, "Limiting Current Densities in Electron Beams," *Jour. Appl. Phys.*, Vol. 10, pp. 715-723, October, 1939.

of an elementary physical picture even though Langmuir and Pierce have covered the subject quite thoroughly. Their work shows that the maximum current density that can be obtained in a line focus is approximately

$$j_{\max} \cong j_o \frac{2}{\pi^{1/2}} \left(\frac{Ve}{kT}\right)^{1/2} \sin\theta \tag{1}$$

where j_o = cathode current density

V = beam potential in volts

$$e/kT = \frac{\text{electronic charge}}{\text{Boltzmann constant} \times T} = \frac{11{,}600}{T} \text{ volts}^{-1}$$

T = temperature in degrees Kelvin

θ = angle of convergence of beam

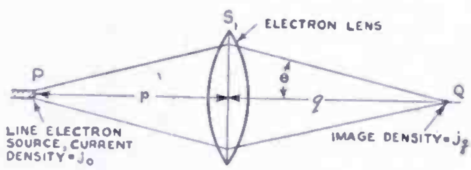


Fig. 1—Simplified picture of an electron-optical system.

Pierce's analysis shows that this theoretical maximum density requires infinite demagnification. He also shows that this expression is in error less than 10 per cent for values of $V > 0.5$ volt.

A good physical picture of the relation can be had by the use of elementary electron optics referring to Figure 1. Consider a source of electrons, P, such as a line cathode, perpendicular to plane of figure, emitting electrons which are focused at an image Q by a cylindrical lens S. If the current density of the object, P, is j_o then in a perfect optical system, in which all the electrons leaving P reach Q, the current density at Q will be

$$j_q = \frac{j_o}{M} \tag{2}$$

where M is the magnification ratio which is roughly equal to $\frac{q}{p}$. It might seem from this elementary picture that the current density at Q could be made as large as desired by making the magnification smaller and smaller by increasing p. The electrons, however, due to their thermal velocities will spread out in every direction from the cathode and in any finite system the current reaching the image will be limited by the lens, by some aperture stop, or, in the case of beam-deflection tubes, possibly by the deflection plates themselves. Thus, as the object distance p is increased the beam will spread out more and more and a larger fraction of the electrons will be caught at S_1, and fewer electrons will get through to the image Q, thus compensating for the increased demagnification. As a result, after increasing p beyond a certain point, the current density j_o might be expected to remain constant. A fairly good quantitative picture can be had by approximating the true emission condition by assuming that electrons are emitted from the cathode uniformly in all directions with a kinetic energy corresponding to a voltage $V_o = \dfrac{kT}{e}$.

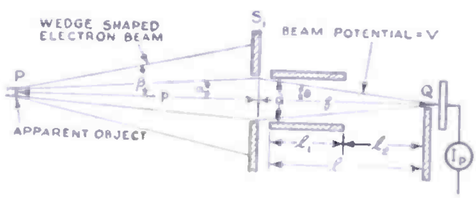


Fig. 2—Cross-sectional view of a beam-deflection tube showing reasons for loss of current density due to a lens stop and finite deflection plate spacing.

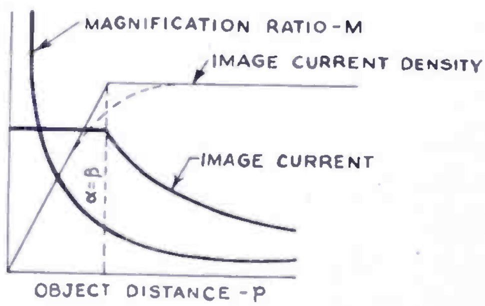


Fig. 3—Effect of varying object distance for an idealized tube.

This means the electrons will spread out in a wedge-shaped homogeneous beam, and the half-angle of spread β will be nearly equal to the ratio of initial velccity to the beam velocity or

$$\beta = \left(\frac{V_o}{V}\right)^{1/2} = \left(\frac{kT}{Ve}\right)^{1/2} \tag{3}$$

Referring to a simple beam-deflection tube as shown in Figure 2, the beam spread half-angle, β, may be greater or less than the entrance angle, α, which is determined by the object distance, p, and the aperture width, d. When β is less than α, all of the current gets through to Q and

$$j_a = \frac{j_o}{M} = \frac{j_o\, p}{q} \tag{4}$$

If p is increased to the point where β just equals α, the current density will reach the maximum value and at this point $\frac{p}{q} = \frac{\theta}{\beta}$, so

$$j_{\max} = \frac{j_0\,\theta}{\beta} = j_0 \left(\frac{Ve}{kT}\right)^{1/2} \theta \tag{5}$$

This expression differs from the more exact approximation, (1), only by a factor $2/\pi^{1/2}$ and that θ replaces $\sin\theta$. Following the elementary picture a little farther it is seen that, as the object distance is increased indefinitely, the fraction of current intercepted will just compensate for the magnification and the image current density will remain at a constant level while the beam current passing thru the deflection plates grows smaller and smaller. This is illustrated in Figure 3 which shows current, current density, and magnification ratio as a function of object distance.

To understand better the role of magnification factor it is well to look at the more general expression for current density in a line focus given by J. R. Pierce[10].

$$j = \frac{j_0}{M} \operatorname{erf} \left[\frac{M^2 \left(\frac{Ve}{kT}\right) \sin^2\theta}{1 - M^2 \sin^2\theta} \right]^{1/2} \tag{6}$$

which assumes only that $Ve/kT >> 1$, as in most practical cases. This expression, when expanded for very small values of $M \sin\theta$, gives

$$j = j_0 \frac{2}{\pi^{1/2}} \left(\frac{Ve}{kT}\right)^{1/2} \sin\theta \left[1 - \frac{(M\sin\theta)^2 \left(\frac{Ve}{kT}\right)}{3} + \frac{(M\sin\theta)^4 \left(\frac{Ve}{kT}\right)^2}{10} \cdots\cdots \right] \tag{7}$$

which reduces to Equation (1) for the case of $M = 0$. Inspection of Equation (7) shows that it is not necessary to go to extremely small values of M to approach the maximum when θ is small, as is usually the case. For example, about 75 per cent of the ultimate current density will be obtained if $$(M \sin \theta)^2 \left(\frac{Ve}{kT}\right) = 1$$

or
$$M \leqq \left(\frac{kT}{Ve}\right)^{1/2} \frac{1}{\sin \theta} \qquad (8)$$

For a practical beam-deflection tube as an example $V = 300$ volts, $\frac{e}{kT} = 10$, $\sin \theta \approx \theta = \frac{1}{55}$, $M = 1$, or a one-to-one magnification ratio is sufficiently small to give 75 per cent of the maximum current density.

Although decreasing the magnification ratio beyond the above value does not materially increase the current density, it does reduce the total current in the image as was shown in the elementary picture of Figure 3. Except for aberrations, the current can be made as small as we please without affecting current density by making the magnification smaller and smaller. This indicates the possibility of using very low beam currents without loss of transconductance, which is an important consideration in connection with input loading and noise.

The total current reaching the image, Q, will depend on the total cathode current $I_c = j_o W_c h_c$ as defined below, and the amount intercepted at the stop, S_1 which determines the entrance angle α. An expression for this current can be derived from the density formula, recalling that $M \cong \frac{\alpha}{\theta}$.

$$I_{\text{image}} = j_o W_c h_c \frac{2}{\pi^{1/2}} \alpha \left(\frac{Ve}{kT}\right)^{1/2} \left[1 - \frac{\alpha^2 \left(\frac{Ve}{kT}\right)^{1/2}}{3} + \cdots\cdots\right] \qquad (9)$$

$$= I_c \frac{2}{\pi^{1/2}} \alpha \left(\frac{Ve}{kT}\right)^{1/2} \left[1 - \frac{\alpha^2 \left(\frac{Ve}{kT}\right)^{1/2}}{3} + \cdots\cdots\right] \qquad (10)$$

where W_c and h_c are width and height of cathode.

Maximum Low-Frequency Transconductance

The maximum transconductance obtainable from an electrostatically focused beam-deflection gun at low frequencies, where transit time is short compared to one period of the deflecting potential, can be derived by combining the above expression for maximum current density with the expression for deflection sensitivity. Consider a rectangular focused beam which is deflected by a pair of deflection plates past an intercepting edge, as illustrated in Figures 2 and 4. If the beam has a current density j_1 at the intercepting edge, then when the beam is deflected a distance $\triangle x$, the change in output current will be

$$\triangle I_p = j_1 W \triangle x$$

where W is the beam width as shown in Figure 4. If the deflection $\triangle x$ is a result of an applied incremental voltage $\triangle E$ to the deflection plates, then, dividing the equation by $\triangle E$ gives $\dfrac{\triangle I_b}{\triangle E} = j_1 W \dfrac{\triangle x}{\triangle E}$.

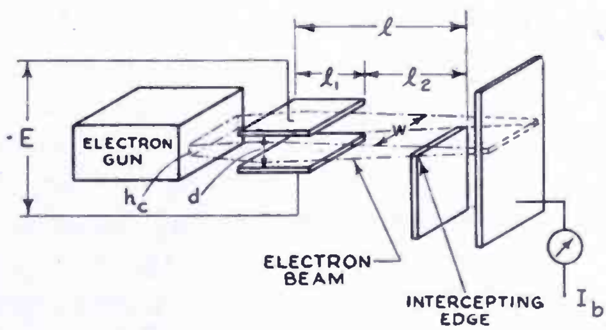


Fig. 4—Illustrative beam-deflection tube.

In the limit, where the changes are small, $\triangle I/\triangle E$ is the transconductance, g_m, and $\triangle x/\triangle E$ is the deflection sensitivity, S.

Thus, the transconductance is

$$g_m = j_1 W S \tag{11}$$

As was seen before, the maximum value of current density, j_1, in the line focus will be,

$$j_{max} = j_o \frac{2}{\pi^{1/2}} \left(\frac{Ve}{kT} \right)^{1/2} \sin \theta \tag{12}$$

Now, in a beam-deflection tube, the maximum angle of convergence, θ, is determined by the deflection plate spacing and the distance, l, from the entrance to the deflection plates to the intercepting edge.

For small values of θ, $\sin \theta \cong \frac{d}{2l}$ so,

$$j_{max} = j_o \frac{1}{\pi^{1/2}} \left(\frac{Ve}{kT} \right)^{1/2} \frac{d}{l} \tag{13}$$

(In this analysis, it is assumed that the potential along the beam is constant and that there is no focusing after deflection.)

The deflection sensitivity at low frequencies is given by the standard formula

$$S = \frac{1}{4d\,V} (l_1{}^2 + 2l_1 l_2) \tag{14}$$

where l_1 = length of deflection plates

l_2 = distance from end of deflection plates to the intercepting edge

For a given beam length, it is evident that the maximum low-frequency deflection sensitivity will be obtained when the deflection plates extend the whole length of the focused beam in which case,

$$S_{max} = \frac{l^2}{4d\,V} \tag{15}$$

Combining Equations (13) and (15), the upper limit of low-frequency transconductance is,

$$g_{max} = j_o \frac{1}{4\pi^{1/2}} \left(\frac{e}{kT} \right)^{1/2} \frac{lW}{V^{1/2}} \text{ mhos} \tag{16}$$

Several interesting facts appear from this expression.

1. The maximum obtainable transconductance is independent of the deflection plate spacing. This suggests that a deflection gun may be built with large spacing and low capacity.

2. The maximum transconductance is proportional to cathode current density and is *independent* of *the total beam current.* This suggests again the possibility of high transconductance to current ratios.

3. Transconductance is proportional to the transit time through the deflection-plate space. Theoretically, therefore, there is no limit to transconductance as the transit angle is made longer. Practically, of course, there are several limits such as mechanical alignment, stray magnetic fields and surface effects.

An example of the ultimate theoretical transconductance for a tube of reasonable dimensions and potentials is as follows:

$$j_o = 0.100 \text{ amperes per square centimeter;}$$

$$T = 1160 \text{ degrees Kelvin,} \qquad \frac{e}{kT} = 10;$$

$$l = 3 \text{ centimeters,} \qquad W = 1 \text{ centimeter;}$$

$$V = 100 \text{ volts;} \qquad g_m = 14{,}000 \text{ micromhos.}$$

All the above was done without the consideration of space-charge limitations. This is justifiable since space-charge effects depend on the *total beam current* which theoretically can be made as small as we please without affecting the transconductance. Practically, the beam current will depend on physical size of the cathode, or the size of the first aperture and on the maximum demagnification. All of the analysis and experimental work included in this report was directed toward the low-current case where space-charge effects are small.

J. R. Pierce[11] has given an expression for the limiting transconductance of a deflection tube expressed in terms of the capacitance of the deflection plates. An examination of his expression shows that it is equivalent to Equation (16), although in a somewhat less convenient form for usual design purposes. The introduction of capacitance in the low-frequency case is not so important as in the high-frequency case where bandwidth considerations are important. The maximum ratio of transconductance to capacitance is derived in the next section for the high-frequency case.

The expression for maximum transconductance derived above was for a cylindrical-lens focusing system. It might appear possible to obtain still higher transconductance by the use of a spherical lens system where the beam can be compressed in width as well as thickness. This case is worked out in Appendix I and the value found to be only one-half the value given by Equation (16) above.

Maximum High-Frequency Transconductance

At frequencies where the transit angle through the deflection plates reaches an appreciable fraction of the period, the force on the

[11] J. R. Pierce, "Theoretical Limitation to Transconductance in Certain Types of Vacuum Tubes," *Proc. I.R.E.*, Vol. 31, pp. 657-663, December, 1943.

electron as it moves through the deflecting plates is not constant and the deflection sensitivity will be less than the low-frequency case.[12] As an example, the ratio of high frequency to direct-current sensitivity is shown in Figure 5 as a function of transit angle, θ_1, through the deflection plates for two special cases. In the general case, the deflection sensitivity is made up of two components, one due to displacement in the deflection plate field and the other due to the transverse drift after the electron leaves the deflection plates. When transit angles are small, these components are in phase but, at high frequencies, these components vary in both magnitude and in phase as the transit angles are changed. General expressions for the magnitude of the deflection sensitivity have been given by a number of writers.[13] A convenient form for the present analysis is,

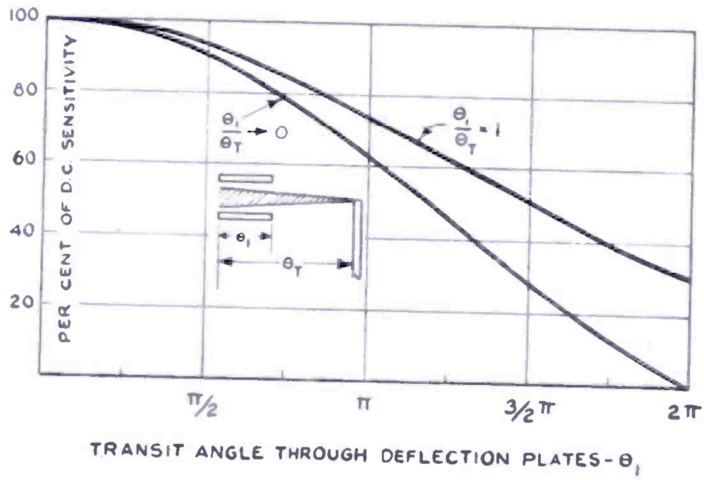


Fig. 5—Loss of deflection sensitivity with increasing frequency. (Curves plotted in terms of transit angle, $\omega\tau$.)

$$S = \frac{1}{\omega^2 d}\left(\frac{e}{m}\right)\sqrt{[(1-\cos\theta_1)+\theta_2\sin\theta_1]^2 + [(\theta_1-\sin\theta_1)+\theta_2(1-\cos\theta_1)]^2}$$

centimeters per volt (17)

where $\theta_1 = \omega\tau_1$ = transit angle through the deflection field,

$\theta_2 = \omega\tau_2$ = transit angle after deflection,

---

[12] J. T. MacGregor-Morris, "Measurements in Electrical Engineering by Means of Cathode Rays," *Jour. Inst. Elec. Eng. (Brit.)*, Vol. 63, p. 1098, 1925.

[13] L. Malter, "Deflection and Impedance of Electron Beams at High Frequencies in the Presence of Magnetic Fields," *RCA REVIEW*, Vol. V, No. 4, pp. 439-454, April, 1941. Also M. R. Gavin and G. W. Warren, "Deflected-Beam Valves for Ultra-High-Frequencies," *G.E.C. Jour.* Vol. 14, pp. 97-103, August, 1946.

$\omega = 2\pi f$ radians per second,

$d =$ deflection plate separation in centimeters,

$\dfrac{e}{m} =$ ratio of electronic charge to mass.

If the transit angle, θ_1, is varied while the total transit angle, θ_T, is held constant, the deflection sensitivity will, in general, go through one or more maxima as shown in Figure 6 where a deflection sensitivity factor is plotted for several values of total transit angle.* It is clear that the maxima are at $\theta_1 = \pi$, 3π, etc. as is found by maxi-

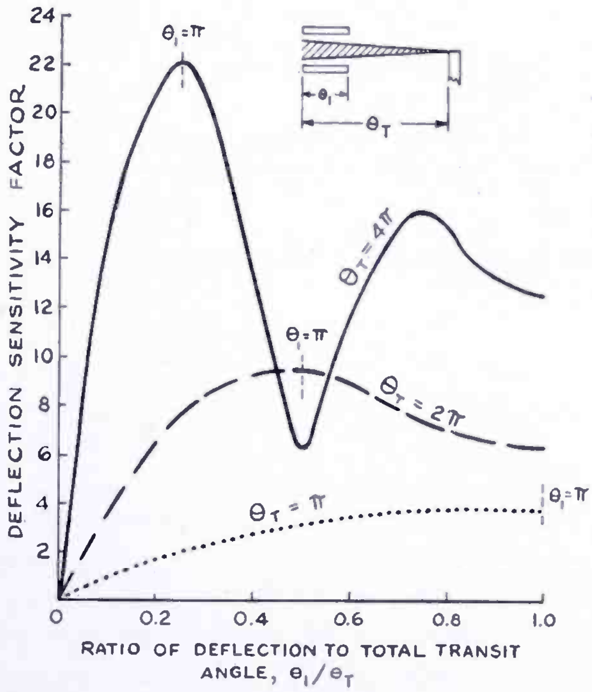


Fig. 6—Curves showing change in deflection sensitivity with length of deflection plates, holding overall beam length (and current density) a constant. (This figure shows maximum transconductance always occurs at a transit angle of π.)

mizing Equation (17). Further inspection shows that in each case the absolute maxima occur for $\theta_1 = \pi$.

Substituting $\theta_1 = \pi$ in Equation (17), the maximum deflection sensitivity is found to be,

$$S_{max} = \frac{1}{\omega^2 d}\left(\frac{e}{m}\right)\sqrt{4 + (\pi + 2\,\theta_2)^2} \qquad \text{centimeters per volt} \qquad (18)$$

* This presentation is due to D. O. North.

It should be noted that even though the high-frequency sensitivity falls off compared to the direct-current sensitivity, that the actual high-frequency sensitivity increases indefinitely as θ_2 is increased. For values of $\theta_2 >> \pi$, the sensitivity varies nearly in proportion to θ_2.

The high-frequency transconductance can now be had by combining Equation (18) with the expression for maximum current density given by Equation (1). Maximum transconductance for a finite transit angle θ_2 is very nearly equal to

$$g_m \cong \frac{j_o W}{\pi^{1/2}} \left(\frac{e}{kT} \right)^{1/2} \frac{V^{1/2}}{\omega^2 l} \left(\frac{e}{m} \right) \sqrt{4 + (\pi + 2\,\theta_2)^2} \qquad (19)$$

where $W =$ the beam width, $V =$ beam voltage,

$l =$ total beam length from entrance to deflection plates to intercepting edge.

This can be put in better form by recognizing that,

$$\frac{\omega l}{\sqrt{\dfrac{2e}{m} V}} = \omega\tau_1 + \omega\tau_2 = \pi + \theta_2$$

then,

$$g_m = \frac{j_o W}{2\omega\pi^{1/2}} \left(\frac{2e}{m} \right)^{1/2} \left(\frac{e}{kT} \right)^{1/2} \left[\frac{\sqrt{4 + (\pi + 2\,\theta_2)^2}}{\theta_2 + \pi} \right] \qquad (20)$$

As θ_2 is increased, the bracketed expression in Equation (20) rapidly approaches the value 2, as shown in Figure 7. For example, for $\theta = 4\pi$ the g_m is up to 90 per cent of the maximum value. The ultimate limit for high-frequency transconductance is therefore given by,

$$\left. g_m \right]_{max} \cong \frac{1}{\pi^{1/2}} \left(\frac{e}{kT} \right)^{1/2} \left(\frac{2e}{m} \right)^{1/2} \frac{j_o W}{\omega}$$

$$\cong 5.4\, j_o \left(\frac{e}{kT} \right)^{1/2} \frac{10^6}{f} W \qquad \text{mhos} \qquad (21)$$

For the usual oxide-coated cathode where T is roughly 1160 degrees Kelvin, the expression for maximum high-frequency transconductance reduces to the simple relation

$$g_m \Big]_{max} = \frac{18}{f_{\text{(in megacycles)}}} j_o W \qquad \text{mhos} \qquad (22)$$

RATIO OF HIGH-FREQUENCY TRANSCONDUCTANCE TO INPUT CAPACITANCE

One of the important figures of merit for wide-band, high-frequency tubes is the ratio of transconductance to capacitance.[14]

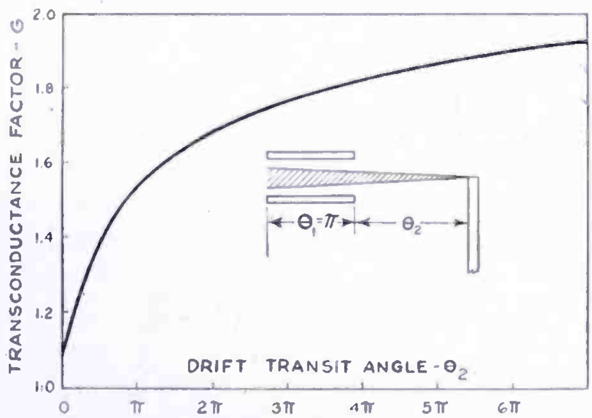


Fig. 7—Curve showing overall transconductance factor as a function of drift distance, assuming optimum deflection plate length.

This ratio is, in fact, proportional to the product of the voltage gain and bandwidth. Having derived an expression for maximum high-frequency transconductance it is interesting to relate this to the capacitance to obtain the bandwidth figure of merit. The case of the rectangular section beam with parallel-plane deflection plates will be considered as it is the one of most practical value. The capacitance between deflection plates will be considered as the only significant one since it is related to the factors affecting transconductance and is likely to be large compared to the output capacitance.

The capacitance between parallel-plane deflection plates is

$$C = a \frac{wl}{d} \times 10^{-12} \text{ farads} \qquad (23)$$

[14] A. P. Kauzmann, "New Television Amplifier Receiving Tubes," *RCA REVIEW*, Vol. III, No. 3, pp. 271-289, January, 1939.

where $l =$ length of plates, $d =$ separation of plates,

$w =$ width of plates $>> d$,

$a =$ a factor depending on the ratio of $\frac{l_1}{d}$

which is 0.0885 for $\frac{l_1}{d} >> 1$.

We have seen in the two preceding sections that the transconductance is apparently independent of the deflection plate separation, and thus it might appear that the capacitance could be made as small as we please by increasing the separation. This cannot be realized because of the effect of fringe fields which were neglected in the earlier work where it was tacitly assumed that $\frac{l_1}{d} >> 1$. In a practical deflection tube at high frequencies, where it is necessary to keep the deflection plates short, the fringe field is a limiting factor. A practical evaluation of this effect leads to a choice of the upper limit of the ratio of spacing to length of about one-half. Using this ratio the capacitance is

$$C_{\min} = 0.28\, w\, 10^{-12} \text{ farads} \qquad (24)$$

It should be noted that this minimum capacitance is *independent of frequency*. The length of deflection plates, of course, will depend upon the frequency and beam voltage in order to maintain the optimum transconductance condition $\theta_1 = \pi$. In going to higher frequencies the length of deflection plates and spacing will decrease together, the capacitance remaining at a constant value given by Equation (24).

The ultimate transconductance-to-capacitance ratio for a high-frequency deflection tube using a rectangular beam is given by combining Equation (24) with Equation (21) assuming the width of beam W equal to the width of the plates, w.

$$\frac{g_m}{C} = 19 \times 10^{18} \left(\frac{e}{kT}\right)^{1/2} \frac{j_o}{f} \text{ mhos per farad} \qquad (25)$$

An upper limit for the voltage gain times bandwidth can now be computed neglecting all capacitances except the deflection-plate capacitance,

$$B_{\max} = \frac{g_m}{2\pi C} \cong 3 \times 10^{18} \left(\frac{e}{kT}\right)^{1/2} \frac{j_o}{f} \qquad (26)$$

It is interesting to work out a practical case to compare with conventional tubes,

$$f = 300 \text{ Mc} = 3 \times 10^8 \text{ cycles/second,}$$

$$j_0 = 0.100 \text{ amperes per square centimeter,}$$

$$\frac{e}{kT} = 10 \ (T = 1160 \text{ degrees Kelvin}),$$

$$B = 3150 \text{ megacycles.}$$

Although this bandwidth merit is an order of magnitude greater than for conventional tubes, it is only fair to point out that this figure could never be reached in practice because of the added capacitance due to the output electrode and leads and because of practical factors limiting the attainment of the maximum transconductance such as will be pointed out in a later section. It should also be noted that this figure of merit decreases inversely with frequency. It is probable that in actual practice that the bandwidth merit of the deflection gun itself may not be much different from that for conventional control. However, because of high transconductance-to-current possibility the deflection gun can be combined with secondary-emission multiplication to give a bandwidth merit much greater than for conventional tubes. Examples of this will be given in next to last section.

Development of a High Transconductance Deflection Gun for High Frequencies

In order that deflection control could be successfully used in high-frequency tubes it was necessary to evolve a simple deflection gun with the following qualities: (1) high transconductance; (2) high ratio of transconductance to capacitance; (3) high ratio of transconductance to plate current and, in some cases, low beam current.

The approach to these objectives was suggested by the analysis presented in the foregoing sections on the limiting transconductance and transconductance-to-capacitance ratio. From such relations it was possible to determine readily some of the design factors such as length of deflection plates, separation of deflection plates and general proportions of the tube without much regard for the type of focusing system used. Computation of these factors will be illustrated before considering the various focusing systems and the choice of a particular scheme. It is not the purpose of this discussion to set up an exact design procedure but rather to indicate in an approximate quantitative fashion the factors in evolving the gun structure which formed the prototype of later guns.

The length of the deflection plates is a good starting point in the design since the transit angle through the plates should be approximately one-half period as shown in section C. The length of plates for a given frequency and voltage is,

$$l_1 = \frac{1}{2f}\left(\frac{2Ve}{m}\right)^{1/2} = \frac{3 \times 10^7}{f} V^{1/2} \tag{27}$$

where $f =$ frequency in cycles per second,
$V =$ beam potential in volts,
$l_1 =$ length of deflection plates in centimeters.

The deflection plate separation, as pointed out in the preceding section, should be increased to about one-half the plate length in order to obtain the best transconductance-to-capacitance ratio (neglecting all other capacitances). Since the transconductance and capacitance are both independent of beam potential V, it is of advantage to use higher potentials to give a larger plate separation thus minimizing surface effects and mechanical difficulties. In receiving tubes it is desirable to keep potentials on the order of 100 volts. Assuming $V = 100$, the dimensions for $f = 3000$ megacycles would be $l_1 = 0.10$ centimeter, $d = 0.05$ centimeter.

It has been shown in Figure 7 that the length of beam from the end of the deflection plates to the intercepting edge need not be more than about three times the length of the plates. Thus, the total length of the deflection system required is about $4l_1$ and the maximum angle of convergence of the beam as shown in Figure 2 is

$$\theta_{\max} = \frac{d}{2l} \cong \frac{1}{16} \tag{28}$$

Since it is desirable to have the beam fill approximately one-half the deflection space, θ reduces to a value of 1/32.

In order to realize an appreciable fraction of the ultimate current density it has been seen that the magnification ratio must be made sufficiently small. The magnification ratio may be determined by substituting in Equation (8):

$$M \leqq \left(\frac{kT}{Ve}\right)^{1/2} \frac{1}{\theta} \cong 1 \tag{29}$$

using values: $V = 100$, $\frac{kT}{e} = \frac{1}{10}$, and $\theta = \frac{1}{32}$.

This means that in order to approach the maximum transconductance the length of the input end of the gun (object distance) should be approximately equal to the length of the deflection end (image distance). Thus, most of the dimensions of the tube are easily determined for a given frequency and given beam voltage. It should be noted that, if very low beam current is also required as well as high transconductance, and if the current cannot be limited sufficiently by the first aperture (one nearest cathode) it may be necessary to increase the object distance beyond that indicated above.

With the approximate dimensions chosen for a simple deflection gun, the type of focusing lens can be considered. A converging lens is desired that with reasonable potential ratios will give the focal length required for proper magnification ratio (roughly unity as shown in Equation (29)). Such a lens is conveniently constructed from a series of aperture lenses with the deflection plates in some cases acting as an aperture lens. Approximate figures for focal length can be computed from the slit aperture lens formula.[15]

$$F = \frac{2V}{E_2 - E_1} \tag{30}$$

where V = aperture potential. E_2 and E_1 are the electric fields on either side of the aperture. Note that $E_2 - E_1$ must be positive to give a converging lens.

There are an endless variety of focusing systems that could be used; some of the more likely forms are shown in Figure 8. In type A, for example, the main focusing lens is formed by the field at the entrance to the deflection plates. This can be treated approximately as an aperture lens where the field inside the deflection plates $E_2 = 0$. Applying Equation (30) it is seen that E_1 must be negative to form a converging lens. This means that V_2 must be less than V_1. It will be noted that the deflection plates are shown much closer to the entrance aperture than to the intercepting edge. This is done to minimize the focusing after deflection which reduces the deflection sensitivity, but this requirement makes it difficult to obtain the desired magnification ratio. A further disadvantage of this simple system is

[15] V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, A. W. Vance, ELECTRON OPTICS AND THE ELECTRON MICROSCOPE, John Wiley and Sons, New York, N. Y., 1945.

that no stopping aperture is provided for limiting the current in the lens region.

Type B also combines focusing and deflection but the magnification ratio in this case is readily controlled by adjusting the distance between the first aperture (nearest to cathode) and the deflection plates. The current can be controlled by the stop S_3 or by an additional stop S_2. It will be shown later that with the double-aperture system formed by aperture S_3 and the deflection plates, a positive lens can be formed with either $V_1 < V_2$ or $V_1 > V_2$. The latter condition is

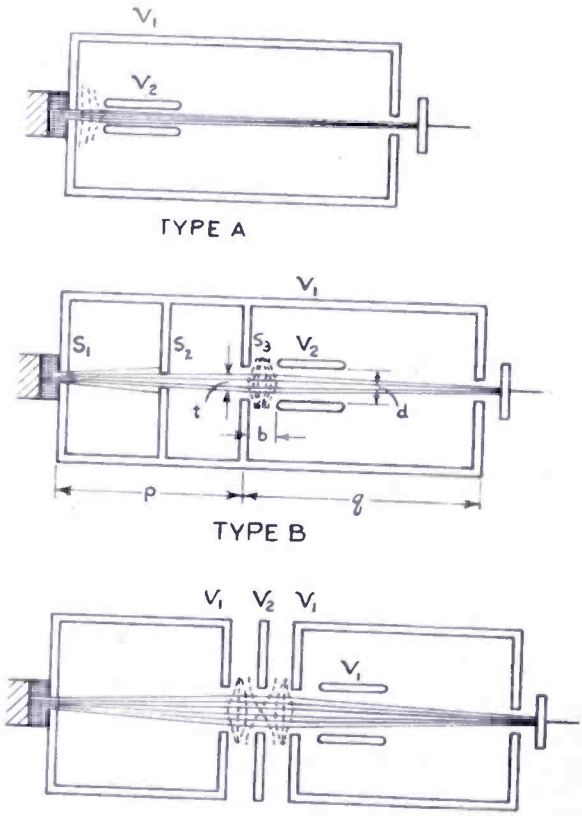


Fig. 8—Three types of electrostatic lens and deflection arrangements for beam-deflection tubes.

usually preferred because it suppresses secondaries from the aperture ahead of the deflection plates. It is this system which has been used in much of the experimental work and it forms the prototype of the deflection guns used in later tubes.

Type C allows the focusing to be controlled independently of the deflection plate potential, a desirable feature when attempting to maximize the transconductance by adjusting transit angle. This type of gun was useful in super-high-frequency tubes using cavity deflection systems. Generally the added complexity of this system has not been found worth while.

Considering the type B lens more closely it is seen that the focal length can be computed approximately by the combination of two aperture lenses to give

$$F \cong \frac{2b}{V_1/V_2 + V_2/V_1 - 2} \tag{31}$$

where b = the effective distance from the aperture to the deflection plates.

A more exact expression can be had by referring to the work of L. H. Bedford.[16]

To obtain the one-to-one magnification ratio indicated for current density considerations, the focal length, F, should be one-half of the image distance, or roughly $2l_1$, in the tentative design discussed above. Assuming further that the separation, b, is equal to the plate separation, $d = \frac{l_1}{2}$, the resulting voltage ratio is $\frac{V_1}{V_2} = 2$ or $1/2$. This was, in fact, the voltage ratio used in much of the early work. In these computations two minor effects were neglected; space charge and the focusing after deflection. The effect of space charge in these tubes is small but it is in the direction to require a stronger lens or a higher ratio of V_1/V_2. In a system such as type B where the potential in the space after the deflection plates is higher than the deflection plates there is a focusing action which tends to decrease the deflection sensitivity. However, in the design considered the field at the exit end of the plates is small compared to that at the entrance end and the focusing effect is generally negligible.

Now with a focusing system determined it is well to consider the electron emission system and to work forward from it to obtain the few remaining factors yet to be determined. It was seen that high cathode current density and low temperature are basic factors in obtaining high transconductance. In fact, as shown in Equation (21) increasing cathode current density is the only clear-cut means for increasing transconductance with a given cathode temperature. In the present work oxide-coated cathodes were used and current densities of from 0.100 to 0.200 amperes per square centimeter were considered reasonable values. To utilize the high current density of the cathode it is necessary to have a sufficient accelerating field in front of the

[16] L. H. Bedford, "Electron Lens Formulas," *Proc. Phys. Soc.*, Vol. 46, pp. 882-888, 1934.

cathode to overcome space charge. This accelerating field must be provided for by an accelerating electrode which may act either as a current limiting aperture or as the first lens in the optical system. In the first case as shown in Figure 9(a) the cathode is large and a slit aperture determines the magnitude of the beam current and forms the object which is imaged on the intercepting edge. In the second case shown in Figure 9(b), the current is determined by a line cathode which is imaged on the intercepting edge. In this case the first aperture acts as a diverging lens which should be sufficiently large to prevent serious aberrations. It should be noted that due to the accelerating field the apparent cathode plane will be at a distance

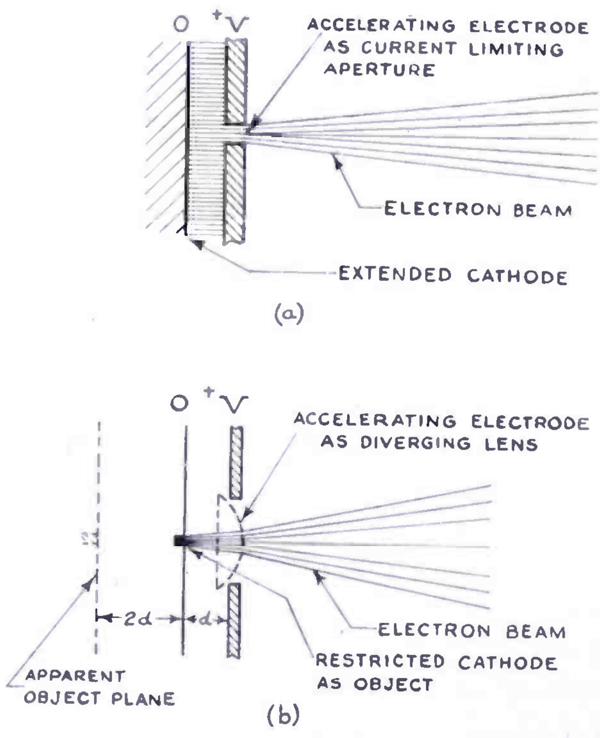


Fig. 9—Two methods of limiting beam current: (a) shows a narrow accelerating aperture; while (b) shows the use of a narrow emitting area on the cathode.

from the accelerator equal to twice the cathode aperture spacing.[15] In the case of full space charge this factor is increased to three. This effect is in most cases small and in any event is in the desired direction of decreasing the magnification ratio.

Both of the cathode arrangements shown in the Figure 9 are workable but it is found that imaging a slit aperture will usually give a sharper-edged image than for most practical cathodes and with the slit aperture it is also easier to limit the beam current to smaller values. In this connection it is well to stress again the point that the transconductance does not depend upon total current and this current can theoretically be made as small as we please. One method of limiting the current is by making the first aperture very small, the other method is to increase the object distance, spreading the beam so that a large fraction of the current is absorbed at a later stop such as S_1. An approximate quantitative expression for the current has already been given in Equation (10). The actual amount of beam spread will be increased beyond that computed for initial velocities because of space charge. This can be readily computed for rectangular beams,[17] and is usually small and again has the effect of decreasing the magnification ratio.

We come now to a consideration of the lens system from the standpoint of aberrations which will reduce the transconductance below the theoretical value and which will have an even greater effect on the maximum ratio of transconductance to plate current than can be attained by working on the very edge of the beam. One important way of reducing aberrations is to restrict the fraction of the focusing lens occupied by the beam, as illustrated in Figure 8(b). The actual transconductance obtainable will thus be reduced from the theoretical transconductance of Equations (16) and (21), by at least the fraction $p = t/d$. In the early tubes to be described this factor was about one-half to one-third. Further reduction might be expected to enhance the transconductance to plate current ratio with some loss of transconductance.

Early Experimental Results

In the early experimental work an effort was made to obtain the highest possible transconductance to current ratio in an electrostatically focused deflection gun. One of the high values which were measured was a ratio of 250 micromhos per microampere, or 20 times more than possible for conventional control methods. Expressed in another way, this corresponds to about a three-to-one change in plate current for input voltage increment of only four millivolts. This result was obtained with a gun similar in design to type B shown in Figure 8 using a 0.002-inch first aperture and 3-centimeter long deflection plates. The accelerating potential was 45 volts and the deflection plate potential 21 volts. The plate current was about 0.01 microampere and the transconductance 2.5 micromhos. Other tubes designed for higher frequencies generally gave ratios of 20 to 30.

Another early deflection gun having a transconductance to current ratio of 20 and an input transconductance of 40 micromhos was used

[17] B. J. Thompson, L. Headrick, "Space-Charge Limitations on the Focus of Electron Beams," *Proc. I.R.E.*, Vol. 28, pp. 318-324, July, 1940.

in combination with a 5-stage secondary-emission multiplier to obtain 100,000 micromhos transconductance at 5 milliamperes plate current. This tube was of the general design shown in Figure 10 except for the number of multiplier stages. The deflection plates were one centimeter long and with accelerator and deflection plate potentials of + 300 volts and + 150 volts. This tube had an input capacitance of 1.5×10^{-12} farads and an output capacitance of 3.5×10^{-12} farads which gives a bandwidth merit at low frequencies of

$$B = \frac{g_m}{2\pi\ (C_{\text{in}} + C_{\text{out}})} = 3000 \text{ megacycles} \tag{33}$$

or on the order of 50 times that of conventional tubes. One of these early tubes used as an amplifier gave a gain of 35 decibels at 450

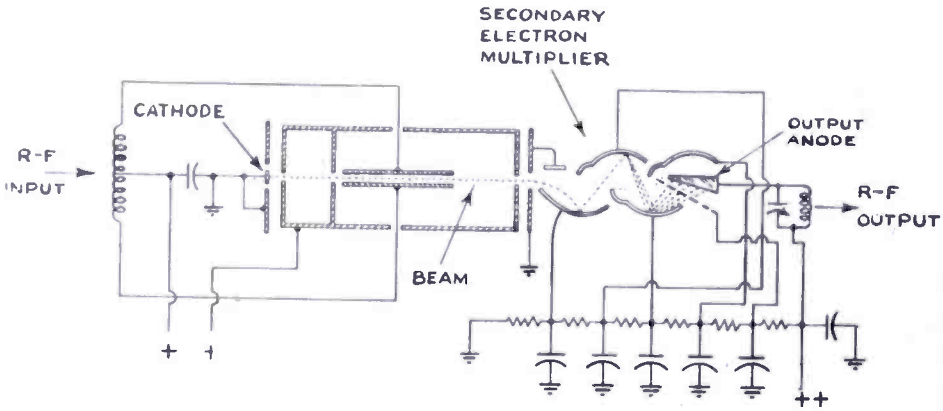


Fig. 10—Cross-sectional view of an early beam-deflection amplifier with 3-stage multipler.

megacycles, despite about 10-to-1 reduction in transconductance due to the combined effects of transit time through the multiplier and reduced deflection sensitivity. In these early tubes the width of beam was usually about one millimeter and the best gun transconductance was about 100 micromhos or 1000 micromhos per centimeter of beam width. In later tubes the beam width was increased to about tenfold without sacrifice of transconductance per unit width. In these earlier tubes the actual transconductance was never much higher than 10 per cent of the theoretical maximum given by Equation (21). One reason for this is that only a fraction of the deflection plate space was utilized. Aberrations and mechanical misalignment probably account for the balance of the reduction. Some idea of the sharpness of the beam can be had from the fact that the total beam thickness at the intercepting edge was 0.001 inch to 0.002 inch wide.

Conclusions

It has been shown that deflection control offers a possibility of obtaining substantial transconductance with low capacitance and low beam currents and with a very high ratio of transconductance to plate current. It has been shown experimentally that useful values of transconductance with low capacitance and low current can be obtained with a simple deflection gun combining focusing and deflection. This type of control is ideally suited for use with a high-gain secondary-emission multiplier to obtain very high transconductance, without excessive capacitance, thus making possible a tube with a bandwidth figure of merit many times greater than for conventional tubes.

The analysis and experimental confirmation of some of the properties of deflection control has already been useful in the development of practical tubes in these laboratories. It may be anticipated that this control method will be useful in other applications, a number of which have already been suggested in the literature. Among these

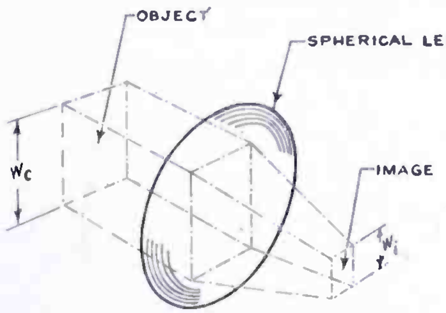


Fig. 11—Representation of electron optics using a spherical lens system to focus both width and thickness of a rectangular electron beam.

are wide-band amplifiers extending into the ultra-high or even the super-high frequency region, detectors and frequency multipliers, and low-frequency or direct-current control applications using the high-voltage sensitivity of deflection control.

Appendix I

The maximum low-frequency transconductance computed in the third section assumed a rectangular beam infinite in extent with focusing by means of a cylindrical lens. If one considers a rectangular beam of finite width as shown in Figure 11 it might appear possible to obtain still higher transconductance by the use of a spherical lens which would allow the beam to be compressed in width as well as in thickness. It will be shown that actually less transconductance per unit width of cathode can be achieved by this method of focusing.

To evaluate the transconductance in this case it is necessary to use the current density expression for the point focus case which Pierce[14] has shown to be

$$j_1 = \frac{j_o}{M^2}\left[1 - (1-\beta^2)\epsilon^{-\beta^2\phi/(1-\beta^2)}\right] \qquad \text{(A-1)}$$

where $\beta = M \sin\theta$,

$\theta =$ angle of convergence of beam,

$\phi = \dfrac{Ve}{kT}$,

$j_o =$ cathode current density,

$j_1 =$ current density at image or intercepting edge.

In a practical deflection tube $\beta << 1$ in which case equation (A-1) reduces to

$$j_1 \cong j_o\,\phi\,\theta^2\left(1 - \frac{M^2\theta^2\phi}{2}\right) \qquad \text{(A-2)}$$

The transconductance is equal to

$$g_m = S\,j_1\,W_1$$

where $W_1 =$ width of the beam at intercepting edge and S is the maximum deflection sensitivty

$$S = \frac{l}{4dV} \qquad \text{(A-3)}$$

Also the maximum angle of convergence is

$$\theta_{max} = \frac{d}{2l} \qquad \text{(A-4)}$$

Combining Equations (A-2), (A-3), and (A-4) and recalling that $W_1 = MW_c$ we obtain an expression for transconductance

$$\frac{g_m}{W_c} = j_o \frac{\phi}{16} \left(1 - \frac{M^2 \theta^2 \phi}{2} \right) \frac{dM}{V} \tag{A-5}$$

Now in the cylindrical lens case it was seen that the transconductance approached a maximum as M approached zero. In the spherical lens case there is obviously an optimum value of M. By maximizing equation (A-5) with respect to M the optimum value is found to be

$$M = \left(\frac{2}{3} \right)^{1/2} \frac{1}{\theta \, \phi^{1/2}} \tag{A-6}$$

(For example if $\theta = 1/32$, $\phi = 1000$, $V = 100$, and $\dfrac{e}{kT} = 10$ $M = 0.8$ or not much different from the practical value of unity assumed in the previous case.) Substituting the maximum value in Equation (A-5) the maximum transconductance per unit width of cathode is

$$\frac{g_m}{W_c} = \frac{(2/3)^{3/2}}{16} \phi^{1/2} \frac{d}{V \theta} \tag{A-7}$$

but, since $\theta_{\max} = \dfrac{d}{2l}$ and $\phi = \dfrac{Ve}{kT}$,

$$\left(\frac{g_m}{W_c} \right)_{\max} = \frac{(2/3)^{3/2}}{8} \left(\frac{e}{kT} \right)^{1/2} \frac{l}{V^{1/2}} = 0.068 \left(\frac{e}{kT} \right)^{1/2} \frac{l}{V^{1/2}} \tag{A-8}$$

mhos/centimeter

When compared with the maximum expression derived in the third section for the cylindrical lens case it is seen that the maximum transconductance per unit width for the spherical lens case is only about one-half.

Grateful acknowledgement is due to E. W. Herold, whose helpful suggestions and encouragement have resulted in the publication of this paper; to D. O. North for his presentation of the deflection sensitivity curves and other helpful suggestions; and to D. B. Langmuir, whose derivations in connection with the line focus case helped to point the way for the developments described in this paper.

MAGNETIC-DEFLECTION CIRCUITS FOR CATHODE-RAY TUBES*

By

Otto H. Schade

Tube Department, RCA Victor Division,
Harrison, N. J.

Summary—*An analysis of the operating cycle in fundamental sawtooth current generating circuits establishes the general requirements for obtaining linear magnetic deflection of cathode-ray beams.*

The graphic representation of the circuit resistance as a load line in the plate characteristics of electron tubes functioning as an electronic switch, furnishes an accurate means of obtaining operating conditions and specifications for the design of practical tubes and circuits.

It is shown that a substantial fraction of the circulating power in certain deflection systems can be recovered as d-c power output from the circuit and, by the use of specific transformation ratios, may be recirculated through the system. Function and design of practical power feedback circuits are analyzed as well as the design requirements for efficient circuit elements, power- and booster-tubes.

Introduction

The trend in the design of magnetically deflected cathode-ray tubes for television uses has greatly increased the magnitude of the field energies required to deflect the electron beam and has made the design of efficient deflecting circuits a major receiver problem. For example, a modern short kinescope having a 50-degree deflection angle and operating at 10 kilovolts would require over 50 watts of direct-current power for full deflection of the beam if circuits such as were used in prewar television receivers should be employed. From the economic viewpoint, this large power requirement would present a rather intolerable condition.

It has long been known that, in principle, an ideal cyclic system for deflecting an electron beam requires only wattless power, and that, in practice, the deflection circuit should function as a power control system and should dissipate only a fraction of the circulated power. The operation of such a control cycle can be illustrated by means of the simple circuit shown in Figure 1(a).[1] In this circuit, closing

---

* Decimal Classification: R138.312 × R583.

[1] Deflection circuits with diode (non-linear)—
(a) A. D. Blumlein, (Britain) U. S. Pat. 2,063,025 (Dec. 8, 1936);
(b) Andrieu, U. S. Pat. 2,139,432 (Dec. 6, 1938).

switch S allows the battery voltage E to drive an exponentially rising current i_1 through the inductance L. A high percentage of the delivered power is stored in the form of increasing magnetic energy in the deflection-coil field. When the current $(+i_1)$ and field strength which give the desired deflection of the cathode-ray beam from the center position have been attained, the switch is opened—Figure 1(b) and (c). The magnetic field must be quickly reversed to give a similar negative deflection (Figure 1(b)). This field reversal takes place without external control because a tuned circuit (LRC) is always formed by L and the inherent capacitance C of the system.

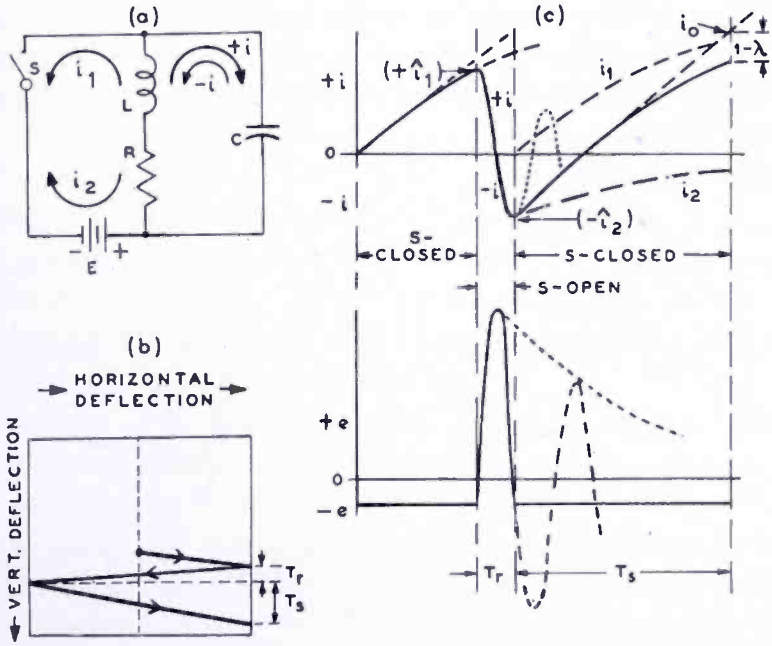


Fig. 1—Fundamental cyclic power-control system for deflecting an electron beam.

Thus, the magnetic energy is converted into potential energy in the electric field of the capacitance by current flow $+i$ from L to C and back again into magnetic energy by reversed current flow $-i$ in substantially one-half cycle of oscillation. The phase relation of voltage and current in low-Q circuits is shown by the vector projections in Figure 2. For values $Q > 3$, the assumption that $\beta = 0$ and $\psi = \pi/2$ is permissible. The retrace time of the beam is, therefore, determined by the constants L and C of the system and can be expressed as

$$T_r \simeq \pi \sqrt{LC} \tag{1}$$

The Q value of the tuned circuit determines the completeness of field reversal; i.e., the peak current ratio which can be written as

$$\hat{i}_2/\hat{i}_1 \simeq \epsilon^{-\pi/2Q} \tag{2}$$

The switch voltage e_s rises to a high value during T_r as shown by the following equation: $$\hat{e}_s = E + \hat{e}_c \simeq E + \hat{\imath}_1 \sqrt{L/C}\, \epsilon^{-\pi/4Q} \quad (3)$$

After reversal of the current flow, the magnetic energy is released for power feedback into the d-c source by closing the switch. The current in the subsequent deflection period T_s may be considered as the sum of two exponential currents $i = i_1 + i_2$ (see Figure 1(c)). The negative current section of the total current i recharges the battery while the beam deflects back to the center position.

Proper functioning of this circuit requires a high L/RT ratio; i.e., incomplete decay of the transient currents. If $R = 0$, the ideal linear current is indicated by i_0, in Figure 1(c). The linearity of the sawtooth current is expressed by $$\lambda = (1 - \epsilon^{-RT_s/L})/RT_s/L \quad (4)$$
This linearity is not sufficient for television purposes with practical

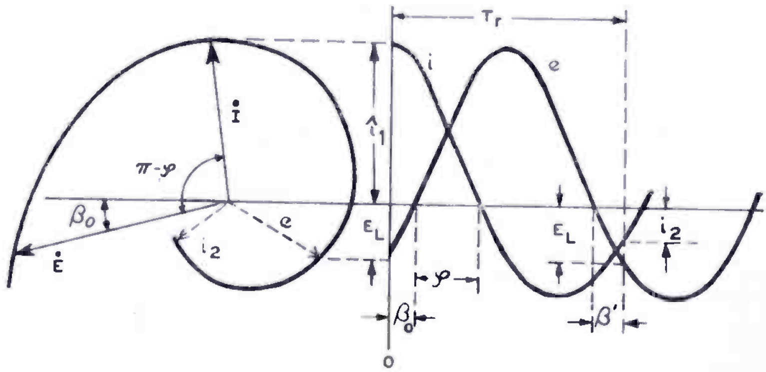


Fig. 2—Phase relations of voltage and current in low-Q circuits during retrace time. ($\tan\phi = 2Q$; $\sin\beta_0 = E_L/\hat{\imath}_1 \sqrt{L/C}$; $\sin\beta' = E_L/\hat{\imath}_2' \sqrt{L/C}$).

coil constants since a deviation of $1 - \lambda = 0.05$ must be considered the upper permissible limit. The necessity for high-speed switching requires the use of electronic switches. The efficiency attainable from such methods has been poor in the past because of the large power losses in the circuit elements and in the switch circuit.

The development of suitable electron tubes for the switch circuit requires that the switching operation be performed at voltages and currents within the range of practical tube characteristics. To provide maximum performance, circuit and tubes must satisfy certain requirements as follows:

1. the deflection current should change linearly with time in circuits operating with present and future television scanning speeds; the circuit Figure 1(a) must, therefore, be modified;

2. the power consumption of the circuit should be minimized by reduction of current or power losses; large energy requirements indi-

cate, therefore, a return of power in a form suitable for feedback or other useful purposes; and

3. the circuit and system must be stable in operation, avoid critical adjustments, and have a moderate cost.

The cyclic transition from oscillatory to aperiodic circuit operation results in unusual operating conditions for the switch tubes. These tubes must control large currents with low voltage drop in the aperiodic phase (T_s), and withstand four to five kilovolts in the oscillatory phase (T_r) of the operating cycle.

The use of a voltage-step-down transformer provides a desirable reduction of the surge voltage on the elements in the secondary circuit and, besides, allows the use of larger capacitances in the secondary circuit, but it increases the current values. Since part of the electronic switching operation can be transferred to the secondary side of such a transformer, its transformation ratio and capacitance effects must be considered in conjunction with tube characteristics.

An analysis will show that linearity of the deflection current cannot be obtained unless the circuit resistance is cancelled during the aperiodic phase of operation. This cancellation requires that the electronic switch act as a negative resistance.

The concept of a negative resistance is very useful in problems of switch and circuit design. It establishes a direct link to general graphic solutions with actual tube characteristics which, in turn, furnish exact numerical values for all operating conditions of circuit and switch. A general analysis will show that certain circuits can be eliminated because of inefficient tube operation and that others can be ignored because of divergent requirements between current control in the tube and linearity in the circuit. The final choice of circuits and of tube types will depend as much on constructional requirements involving high-voltage insulation, low-capacitance mounting of circuit elements, and control of critical tolerances, as it will on operating efficiency and on power requirements.

Deflection Circuits With Controlled Negative Resistance

Linear deflection of the cathode-ray beam requires a deflection current with a constant rate of change with time. A deflection current i with constant rate of change induces a constant inductive voltage

$$e_L = L\, di/dt = \text{constant} \tag{5}$$

This relationship is obviously not true of the circuit of Figure 1, where

$$e_L = E - iR \tag{6}$$

To obtain linear deflection, therefore, it is necessary to eliminate the iR drop in the circuit. Although the deflection-coil resistance R cannot be made zero, a generator modulated in synchronism with the current i can be employed as shown in the circuit of Figure 3, to supply compensation for the iR drop in the aperiodic phase of circuit operation. Such compensation requires that the characteristic of the generator obey the law $\Delta e/\Delta i = -R$

a condition which can be met by utilizing the control characteristic of an electron tube. At first thought, it may appear that this control should be a direct function of the current i, but such control will cause instability in the form of relaxation oscillations because the circuit would then have no time constant of its own within the controllable current range of the tube. The control, therefore, must be effected by a voltage having an independent time constant, i.e.,

$$\Delta ep/\Delta ip = -R = f(e_g), f(t) \tag{7}$$

The quantity $-R$ is a negative resistance and is shown graphically in Figure 4. With the operating path $-R$ given, it is now relatively simple to find the control voltage $e_g = f(t_s)$ required during the trace time T_s as a dependent variable from the electron-tube plate characteristic. It will be shown that circuit modifications are required by the fact that an electron tube can conduct only a positive current. The operating conditions during retrace time (T_r) place additional requirements on the switch circuit and determine the corresponding control voltage $e_g = f(t_r)$.

The complete control voltage wave $e_g = f(t)$ must also remain within certain limits of waveform and amplitude which can be generated in auxiliary circuits with the voltages available.

Electronic Switches With Negative Resistance

Switch with one electron tube

The function of an electron tube in generating a negative resistance is explained by the graphic construction of Figure 5. Stability of operation is obtained by varying the control-grid voltage of the tube as a function of time as required by Equation (7). Figure 5 shows that the operating path $-R$, constructed from current and voltage waves, is the familiar load line in the tube characteristic. The negative control signal voltage $e_g = f(t_s)$ is obtained by plotting the intersections of the grid voltage curves with $-R$ against Δt_s. Current range and control signal obviously depend on where $-R$ lies in the plate characteristics of the tube.

The bidirectional current $\pm i$ in the deflection circuits of Figure 1 or Figure 3 requires an operating path $-R$ passing through the zero

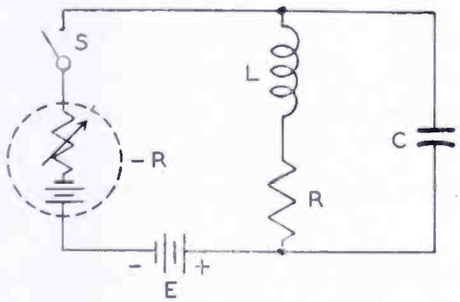


Fig. 3—Control system employing a generator to compensate for *iR* drop in deflection-coil resistance for linear deflection of an electron beam.

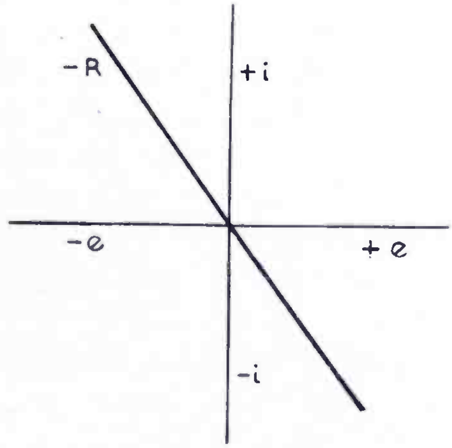


Fig. 4—Graphical representation of equivalent negative resistance generated in circuit of Figure 3.

current value. A single tube with one current-carrying electrode can, therefore, function only when the negative return current $-i$ is eliminated. The tuned circuit must then be made aperiodic by using a shunt resistance to dissipate all of the stored energy. This solution is satisfactory for low-frequency (vertical) deflection circuits where a short retrace time is obtainable with deflection coils of large turn-number and inductance. Equation (5) shows that the inductive volt-

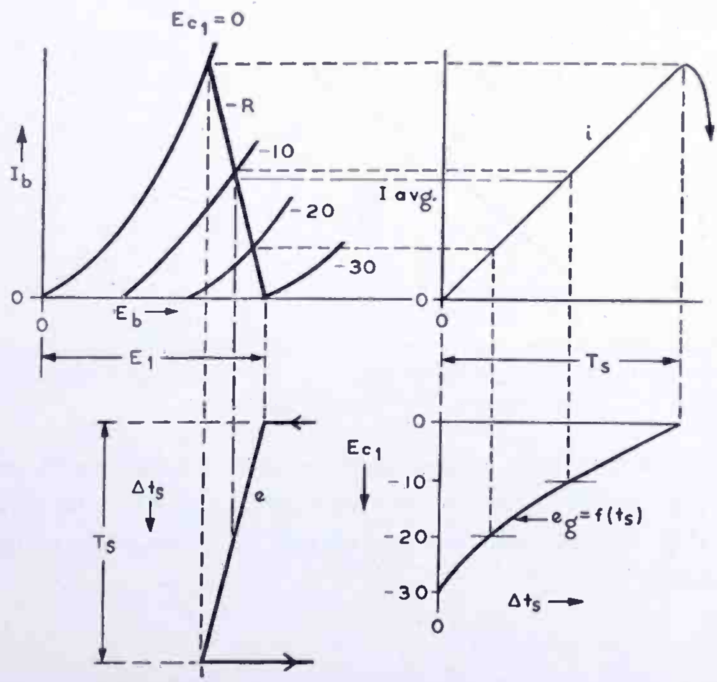


Fig. 5—Graphic construction indicating function of a single electron tube in generating a negative resistance.

age component is small and, therefore, the deflection coil acts predominantly as a resistive load R which can be matched by suitable transformer ratios to the tube characteristics.

This method, however, is too inefficient to be considered for high-frequency (horizontal) kinescope deflection, where dissipation of the stored energy in a damping resistor will double the required input power and lengthen the retrace time. It is, therefore, essential in this case to design the switch circuit for conduction of bidirectional currents.

A bidirectional current can be conducted by two electron tubes

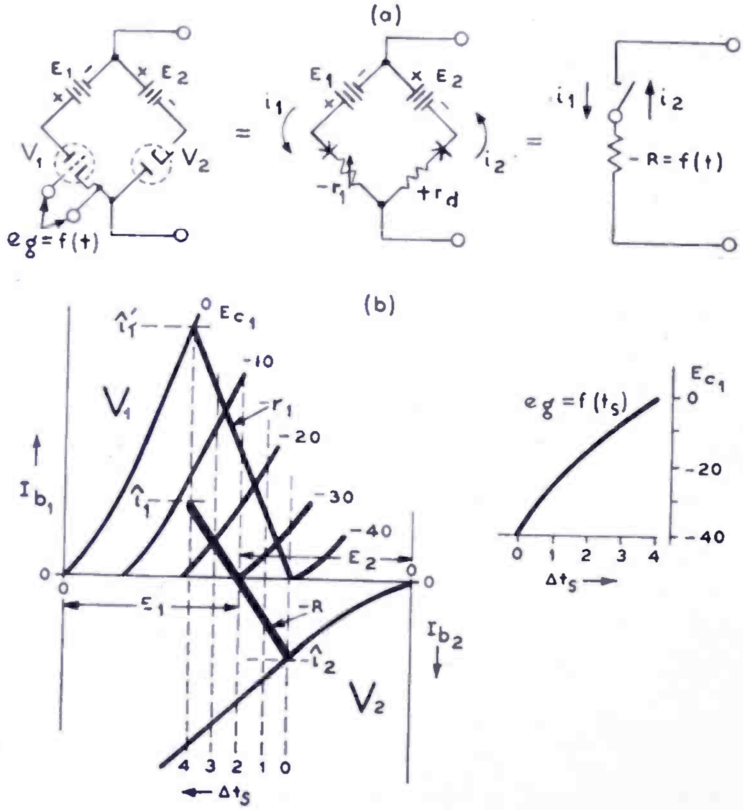


Fig. 6—Bridge and equivalent circuits for obtaining bidirectional currents; and, characteristic curves showing tube operating conditions for balance of steady current component.

forming a bridge circuit. The steady current component required by tube operation in the positive current region can thus be balanced out. It follows that to accomplish this, one of the tubes must have a controllable voltage drop.

Switch with triode and diode

A bridge circuit containing one triode V_1 and one diode V_2 is shown in Figure 6(a). Tube operating conditions for balance of the steady current component are obtained from the joined characteristics shown in Figure 6(b). The load-line slope $-R$ is equal to the known positive resistance R of the deflection coils, ending at $\hat{\imath}_1$ and $\hat{\imath}_2$ in the respective quadrants of V_1 and V_2. The diode characteristic is drawn

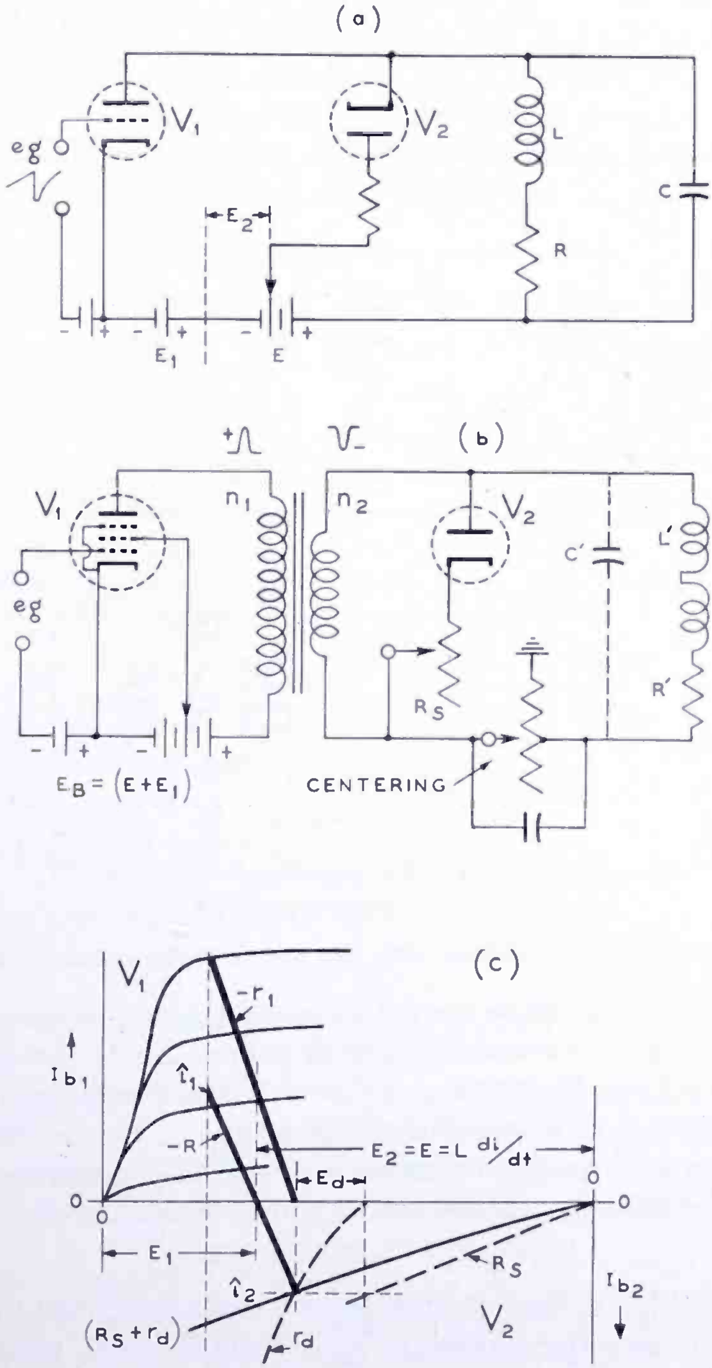


Fig. 7—Rearrangement of bridge circuit; circuit with addition of transformer; and, characteristic curves showing effect of increasing diode resistance by adding R_s.

so that $\hat{\imath}_2$ lies on the diode line. This determines the plate battery voltage E_2. The negative resistance $-r_1$ which must be generated by V_1 is constructed point by point by the addition of the diode current and the current required by $-R$ at respective plate voltages (parallel to the current axis). The characteristic of V_1 is now shifted until $\hat{\imath}_1$ lies at or below the zero grid-voltage line of V_1, thus determining E_1. The control grid voltage $e_g = f(t_s)$ is readily obtainable from the intersections of $-r_1$ with the grid voltage curves E_{c_1} of V_1.

It is apparent from the construction that V_1 must furnish a peak current in excess of the peak-to-peak deflection current. A higher

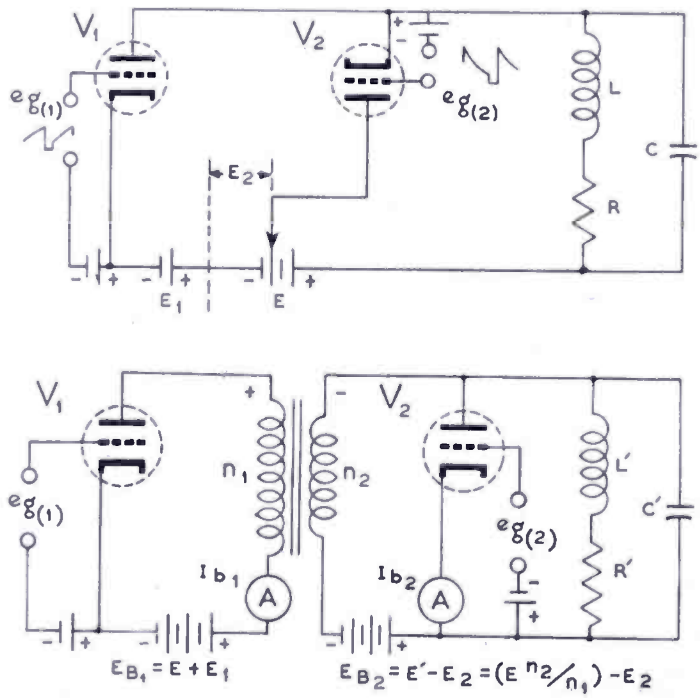


Fig. 8—Practical circuits with two controlled electron tubes.

diode resistance will reduce the peak current but will require a larger voltage E_2. Rearrangement of batteries results in the circuit Figure 7(a) where E_2 can be raised to equal E without additional voltage. With the addition of a transformer, this circuit (Figure 7(b)) will be recognized as a prewar deflection circuit. The diode resistance is adjusted by adding a variable series resistance R_s; linearity occurs when

$$R_s = [E - (\hat{\imath}_1 R + E_d)]/\hat{\imath}_2 \tag{8}$$

and when the grid-voltage wave of correct shape is applied to V_1 (see Figure 7(c)). The transformation ratio from V_2 to V_1 is taken into account by multiplying current and voltage scales of V_1 with the turns ratios n_2/n_1 and n_1/n_2, respectively. Although giving linearity, the circuit has a low deflection-current efficiency, similar to that of an aperiodic circuit, because of the large circulating bridge-current. It has, however, a shorter retrace period.

Switches with two controlled electron tubes

Considerably better performance is obtained with bridge circuits containing two controlled electron tubes.[2] Each tube contributes to the

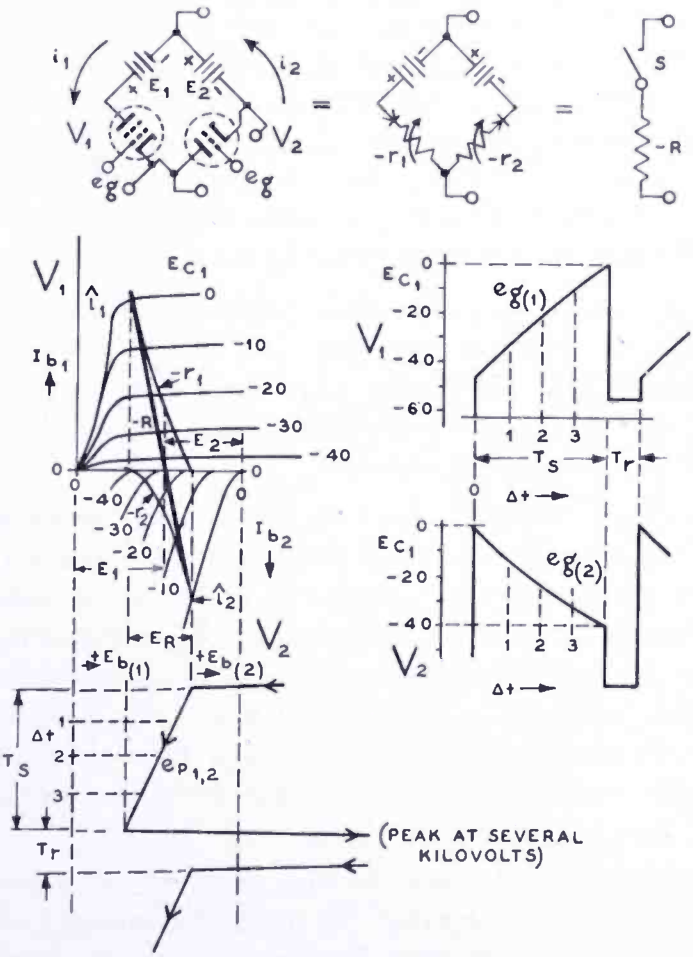


Fig. 9—Equivalent bridge circuits for Figure 8 and characteristic curves for the two tubes.

negative resistance $-R$ and circulating bridge-currents can be reduced to small values. The rearranged circuits are shown in Figure 8.

Switch circuit and combined characteristic of two controlled tubes V_1 and V_2 are illustrated in Figure 9, and may be treated like the

---

[2] Circuits with negative resistance—
(a) Max Geiger, U. S. Pat. 2,225,300 (Dec. 17, 1940); filed 6/9/38.
(b) R. C. More, U. S. Pat. 2,251,851 (Aug. 5, 1940); filed 6/16/39.
(c) W. A. Tolson, U. S. Pat. 2,280,733 (April 21, 1942); filed 6/30/39.
(d) H. A. Wheeler, U. S. Pat. 2,235,131 (March 18, 1941); filed 10/25/39.
(e) Otto H. Schade, U. S. Pat. 2,382,822 (Aug. 14, 1945), filed 6/30/42.

familiar push-pull amplifier characteristic with respect to graphic addition of currents. It is, however, important to keep in mind that voltages and currents are non-sinusoidal and asymetric and that the current change di/dt along the load line $-R$ must be constant.

The starting point for the graphical construction of Figure 9 is again the line $-R$ with end-points at $\hat{\imath}_1$ and $\hat{\imath}_2$ given by the required deflection and the Q value of the tuned circuit LCR. The characteristic of V_2 is drawn so that $\hat{\imath}_2$ is located on the zero-bias curve, determining, therefore, the plate voltage E_2. The plate voltage E_1 for V_1 is obtained in a similar manner. Good efficiency and uncritical matching indicate grid-voltage amplitudes which cause cutoff on one tube when the other tube carries peak current.

The constant summation load characteristic $-R$ can be obtained with many different pairs of individual load characteristics $-r_1$ and $-r_2$ (see Figure 10). Each pair requires a specific pair of grid voltages $e_{g(1)}$ and $e_{g(2)}$. Given one grid voltage wave $e_{g(1)}$, the other voltage $e_{g(2)}$ is determined as follows:

Divide the plate-voltage change E_R into equal increments, $\triangle E$, occurring at equally-spaced time increments, $\triangle t$. Locate the corresponding grid voltages on $e_{g(1)}$ (Figure 9) by using the same number of time increments, $\triangle t$. The intersections of these grid-voltage curves in the V_1 characteristic with corresponding $\triangle E$ values furnish, therefore, the desired currents i_1 and the load path $-r_1$. Obtain the load path $-r_2$ by subtracting the current values i_1 from the total current in $-R$. Plot the waveform $e_{g(2)}$ against time from the intersections of plate voltage, grid voltage and load path $-r_2$.

The waveform and amplitude of the grid voltage determine the magnitude of the matching current (in the bridge circuit), which contributes nothing to the deflection current. Figure 10 shows various degrees of current efficiency. Figure 10(a) indicates ideal but very critical class B operation for zero matching current and for an ideal tuned circuit having $R=0$. For this arrangement, $\hat{\imath}_1$ equals $\hat{\imath}_2$. This condition represents ideal current utilization and has the significant relation $\hat{\imath}_1 + \hat{\imath}_2 = 8\, I_{b1}$, where I_{b1} is the average current value in V_1. For this condition, the current utilization factor or current efficiency becomes

$$\eta i = (\hat{\imath}_1 + \hat{\imath}_2)/8\, I_{b1} \tag{9}$$

A maximum current efficiency of ηi equal to unity does not indicate that the power loss in the electron-tube bridge circuit is zero; instead, it indicates that the power loss in the circuit LRC is zero and that there is a minimum current drain from the supply voltage source.

A good stable operating condition obtainable with practical grid-voltage waveforms and $\eta i = 0.63$ is shown by Figure 10(b). Figures 10(c) and 10(d) show conditions arising from poor adjustments and low-Q circuits. Figure 10(e) shows the comparatively low current-efficiency $\eta i = 0.25$ of the circuit Figure 7 when an uncontrolled diode is used to obtain a linear sawtooth current, while Figure 10(f) illustrates the gain in deflection obtainable with the same circuit by allowing a non-linear current i. This circuit requires a *positive* load line slope in the construction Figure 7(c) in order to effect diode cutoff.

Switch with pentode and plate-voltage-controlled diode

The characteristic of the pentode-triode switch (Figure 9) can be

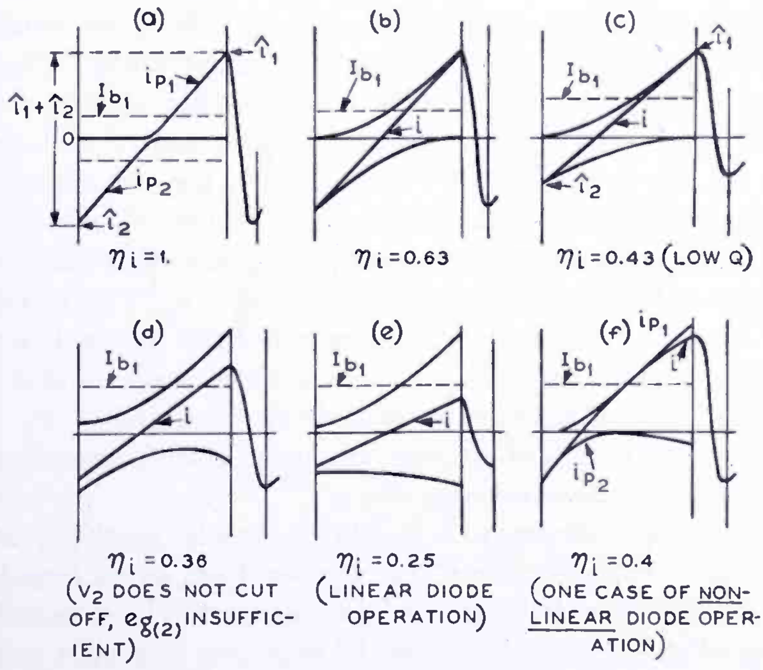


Fig. 10—Degrees of current efficiency with varying grid voltages, circuit adjustments, and tubes. $(i = (\hat{i}_1 + \hat{i}_2)/8\, I_{b_1})$.

duplicated with pentode and diode by displacing the diode line in the composite characteristic (compare Figure 6(b)) towards the left as a function of time, thus forming in effect a triode characteristic. This displacement is accomplished, in principle, by the insertion of an auxiliary synchronous generator (E_z) in series with the diode as illustrated in Figure 11. Efficient operation requires diode current cutoff by a decreasing diode plate voltage e_d, while linearity of deflection current requires an increasing voltage $E_L + iR$ across the deflection coil system. The auxiliary generator voltage is, therefore, specified by

$$e_z = f(t_s) = -(e_d + e_R) \qquad (10)$$

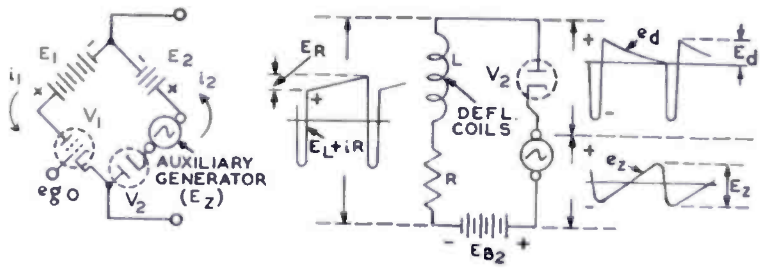


Fig. 11—Diode control circuit and equivalent switch current.

with the peak-to-peak value
$$E_z = -(E_d + E_R) \tag{10a}$$
To duplicate the triode action, the waveform of e_z should approximate a sawtooth. Insertion of e_z into the diode circuit can be effected by means of an auxiliary transformer T_2, having one of its windings in series with the plate or cathode connection of the diode. The control voltage can be supplied by an auxiliary small power tube V_3 or by the power tube V_1 as indicated in the fundamental switch circuits given in Figure 12. In the circuit of Figure 12(b), the transformer T_2 is given a small step-down ratio because the current i_1 must not only supply shunt losses but must also counteract and exceed the diode current i_2 in the secondary current $i_s = (n_p/n_d)\, i_1 + i_2$. The voltage e_z is developed across an impedance Z representing the internal impedance of the auxiliary generator and may be connected across either or both of the closely coupled primary or secondary windings of T_2. Proper phasing of e_z in these and in most practical circuits requires phase reversal of the secondary voltage.

For best circuit efficiency Z should be purely reactive. In order to obtain a good sawtooth waveform, Z should act as an impedance to at least several harmonic components contained in the sawtooth wave.

Although the simplest impedance of this type is a pure resistance, its use, however, results in some power loss. It is, therefore, of

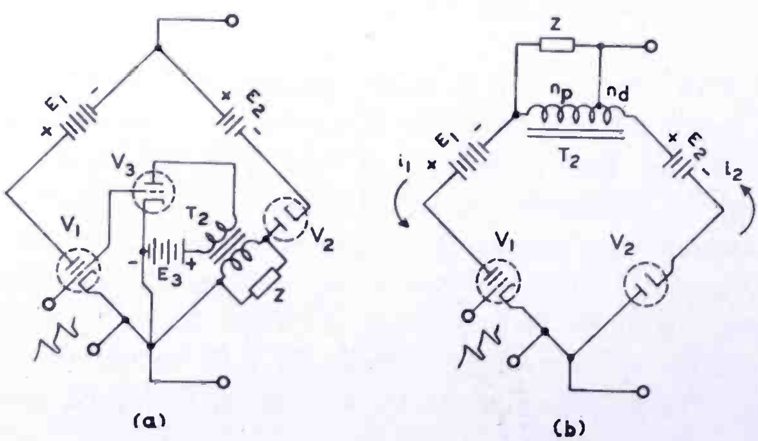


Fig. 12—Diode switch circuits with negative resistance (—R).

interest to determine the limits within which amplitude, waveform, and phase of e_z may be varied for practical operating conditions of the electron-tube switch and linear deflection current. The waveform will indicate the requirements for the impedance Z with respect to higher-order harmonic components.

The graphic solution for the control voltage e_z is shown in Figure

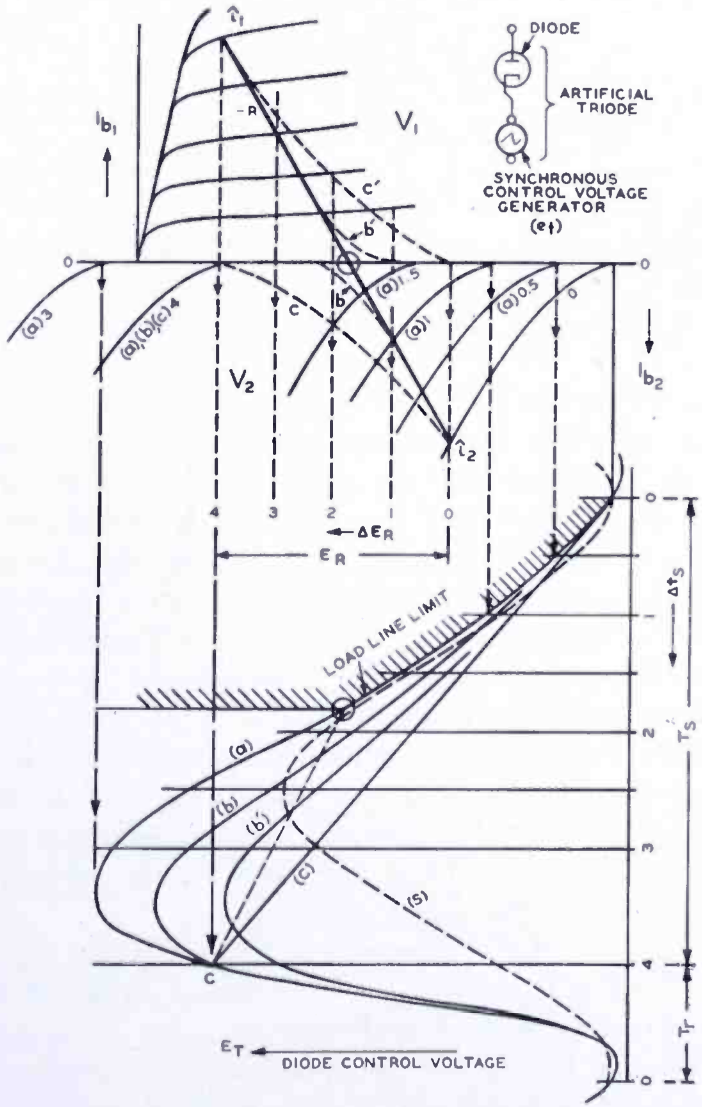


Fig. 13—Characteristics and control voltages for linear deflection with controlled diodes.

13. The limiting class B case (see Figure 10(a)) requires that the individual load path of V_2 should follow the load line $-R$ to the cut-off point 0 at $i=0$ where it turns sharply to the left and remains at $i=0$ for the second half of the load cycle. The current of V_1 is zero in the first half of the load cycle, its load path turning sharply at point 0 to follow the load line $-R$ to $\hat{\imath}_1$.

The diode* characteristic is drawn through the initial current $\hat{\imath}_2$. The cutoff point places the value of the control voltage at $T_s = 0$. The voltage E_R changes linearly and is divided into equal increments $\triangle E_R$, which correspond to equal time increments $\triangle t_s$ on the control-voltage time base. The instantaneous value of e_T at a time $\triangle t_s = 1$ is found as the voltage increment by which the diode line must be shifted to the left to intersect $-R$ at the corresponding plate-voltage change $\triangle E_R = 1$. In this manner the control-voltage wave (a) is obtained, indicating the load-line limit as far as point 0 ($\triangle T_s = 1.8$). The cutoff point moves, then, along the broken line to point C. The waveform of e_z in the range $\triangle t_s = 1.8$ to $\triangle t_s = 4$ (points 0 to C) is unimportant so long as it remains to the left of the broken line. It may thus have the practical form indicated by curve (a) which, for the example, represents the closest approach to a sine wave giving efficient linear deflection. It is obvious that waveform, phase, and amplitude are critical.

It is not difficult to prove that practical operating conditions require the control-voltage wave to remain to the right of the shaded area marked by the load-line limit up to the crossover point 0 and that once having crossed the cutoff line (dashed) it should remain on its left side for the remaining portion of T_s. Waveforms (b) and (c) fill this requirement while the sine wave (s) crosses the cutoff limit again at $\triangle t_s = 2.7$ thus causing a premature return of the diode current and requiring, in turn, an enormous rise in power-tube plate current i, if deflection linearity is to be maintained.

Curve (b′) shows that an amplitude reduction of (b) causes a second crossover indicating that the magnitude of Z as well as the phase of e_z must be carefully adjusted. Voltages approaching the sawtooth shape (c) are less critical and more readily controlled as the impedance of Z approaches a pure resistance.

A practical form of the impedance Z which permits adjustment of phase and waveform is shown in the rearranged circuit of Figure 14.

The inductance L_2 of the transformer T_2 forms a damped resonant circuit with C and the adjustable resistance R_2. Variation of C (or L_2) provides a phase adjustment for E_z, while R_2 controls waveform and amplitude. Because of this interaction, the constants L_2 and C as well as the stepdown ratio must be properly selected. Higher impedance LC circuits and larger step-down ratios permit the use of larger values of resistance resulting in more flexibility than is obtainable with low-impedance circuits and transformers with smaller step-down ratios which require a higher value of Q to build up sufficient control voltage.

* Diode can be gas or vacuum type.

It is apparent that the plate supply voltage E_{B_1} must be increased to include the voltage drop across Z. This increase is given by $\triangle E_{B_1} = E_{Z(x)}/2$ when Z is purely reactive and increasing towards $\triangle E_{B_1} = E_Z = \hat{\imath}_{p1} R_2$ when Z is a pure resistance.

The reactive voltage-drop and the corresponding power are substantially regained by the booster tube V_2 in the form of increased dc power output into the battery $E_{B_2} = (E' - E_2) + (E_{Z(x)}/2)$ (See section discussing power-feedback circuits — page 524.)

Although similar in deflection efficiency, the controlled diode circuit has more interacting parameters than a triode booster circuit which by its natural triode plate characteristic facilitates accurate matching of the V_1 and V_2 characteristics.

Switch-Circuit Properties Required by the Oscillatory Phase

Since the electron tube bridge must be an open circuit during retrace time T_r, a rapid plate-current cutoff in V_1 and V_2 is required

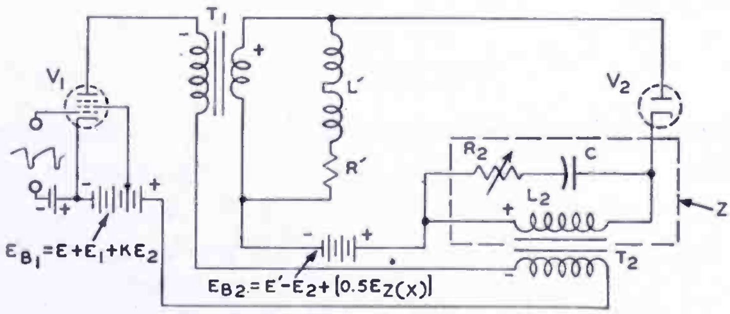


Fig. 14—Rearranged deflection circuit with controlled diode (as Figure 12 (b)) combining good efficiency and linear deflection.

to prevent damping of the tuned circuit. The switch voltage e_s in Figure 1 is obviously the sum of the battery voltage E and the voltages developed by the current i flowing through R and L. Thus,

$$e_s = E_B + (iR + L\,di/dt) \tag{11}$$

The second term in Equation (11) is the plate load voltage of the tubes V_1 and V_2. It varies with time as shown in Figure 9. In the oscillatory time, T_r, the voltage $L\,di/dt$ has a sine-wave shape and rises to a high peak value, thus,

$$\hat{e}_L \simeq \hat{\imath}_1 \sqrt{L/C}\,\epsilon^{-\pi/4Q} \tag{11a}$$

In the aperiodic time T_s, the voltage $L\,di/dt$ has the substantially constant value

$$E_L = -L\,(\hat{\imath}_1 - \hat{\imath}_2)/T_s = -L\,\hat{\imath}_1\,(1 + \epsilon^{-\pi/2Q})/T_s \tag{12}$$

since the exponential terms approach unity in high-Q circuits. When $Q \simeq 6$, $\hat{\imath}_2 \simeq 0.85\,\hat{\imath}_1$, and the Equations (1), (11), and (12) furnish the approximate relation

$$\hat{e}_L \approx 1.7\,E_L\,T_s/T_r \tag{13}$$

The maximum switch voltage $\hat{e}_s$ is thus inversely proportional to the retrace time. Current cutoff in a triode in the V_1 position requires, therefore, a highly negative grid voltage pulse during T_r because, as Equation (13) indicates, positive peak voltages of several kilovolts may occur in practical circuits (see Table IV — page 535). A pentode or beam power tube is, therefore, most suitable as power tube V_1 since screen-grid tubes can be cut off with a small negative grid voltage which is substantially independent of the value of the plate voltage.

The "booster" scanning tube V_2 receives a negative plate voltage during the major part of T_r except during the short time angles β_0 and β' (see Figure 2). It can, therefore, be a diode, triode or pentode if its switch function only is considered. The grid voltage should, however, contain a negative pulse of sufficient magnitude to effect rapid

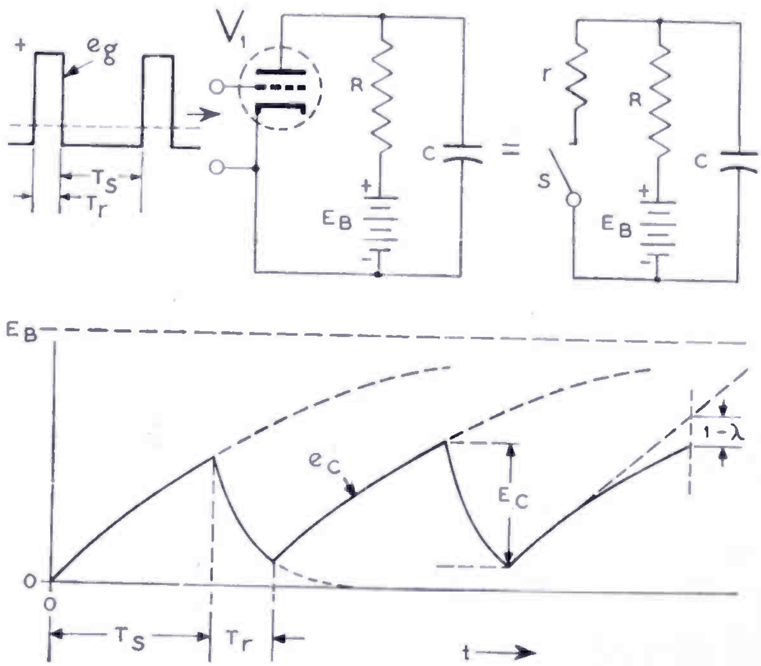


Fig. 15—Generation of control voltage for power tube, V_1.

cutoff within the angles β (Figure 2) where the plate voltage (e_p) is positive (see Figure 9).

Generation of the negative resistance $-R$ during T_s can be accomplished equally well with diodes, triodes or pentodes. However, proper functioning of the deflection circuit also requires that the tuned circuit become aperiodic at the beginning of T_s to prevent further current oscillations. The electron tube switch must, therefore, be a short-circuit or a low value of positive resistance (r_p) to damp out all parasitic voltage or current fluctuation while acting as a controlled negative resistance $-R$ only for the desired deflection current i.

It is thus essential that at least one of the tubes V_1 or V_2 have a low plate resistance (r_p). If V_1 is a high-impedance screen-grid tube, the booster scanning tube V_2 must also act as a damping resistance requiring the low plate resistance obtainable with a low-μ triode or diode.

It is possible to obtain low effective r_p values with high-μ tubes by inverse feedback but difficulties in obtaining properly phase feedback voltages limit the usefulness of this method for practical circuits.

Generation of Control Voltages (e_g)

The control-grid voltage for the power tube V_1 is usually generated in an RC circuit by means of an electron-tube discharge switch. (Figure 15). In this circuit, the capacitor voltage rises exponentially during the charging period T_s. The linearity λ of the voltage rise is expressed by

$$\lambda = (1 - \epsilon^{-T_s/RC})\ RC/T_s \tag{14}$$

The following approximations may be used for values $\lambda > 0.75$ or values of peak-to-peak sawtooth voltage $E_c < 0.5\ E_b$, by assuming substantially complete discharge of C during T_r; i.e., for $r << R$,

$$E_c \approx 2E_b\ (1 - \lambda) \tag{15}$$

and

$$T_s/RC \approx (2/\lambda) - 2 \tag{16}$$

A negative pulse to cut off the power-tube plate current is obtained by addition of a resistance load in series with C. The pulse width should correspond to T_r of the deflection circuit. When shorter, as in case of a blocking-oscillator drive, the cutoff pulse can be obtained by feedback from the transformer secondary (see Figure 17).

The control-grid voltage for the booster triode V_2 can be obtained from the pulse voltage without the use of additional tubes.[3] The circuit function is illustrated by the equivalent of circuit Figure 16 which is a special case of the discharge circuit.

$$\text{With } \epsilon^{-T_r/RC} = K_1 \qquad \text{and } \epsilon^{-T_s/RC} = K_2 \tag{17}$$

the peak-to-peak capacitor voltage is given by

$$E_c = (E_1 + E_2)\,(1 - K_1)/[K_1 + (1 - K_1)\,(1 - K_2)] \tag{18}$$

The integrated voltage, e_c, is relatively small because of the short charging time. The differentiated voltage, e_r, is the pulse voltage minus the capacitor voltage as shown in Figure 16(b). The sawtooth section has the same amplitude as e_c, but opposite polarity; the negative pulse during T_r is more than sufficient for cutoff of V_2. The circuit constants are again given by Equation (16) with λ ranging from 0.3 to 0.8 according to grid-voltage wave shape. The voltage E_2 in Equation (18) is approximated by the average value of the sinusoidal pulse voltage $E_2 \cong 0.63\ e_2$. A practical control-voltage-generating circuit is shown in Figure 17. The impedance of the circuit $R_2\,C_2$ is limited by the tube electrode capacitances to certain maximum values.

It should also be noted that the voltage pulse of practical transformers is not a smooth half-sine-wave pulse. It contains harmonic frequencies due to leakage reactance tuning with varying coupling and phase relations between windings. These ripples are degenerative in V_2 when the negative secondary pulse is used for generation of e_g of V_2.

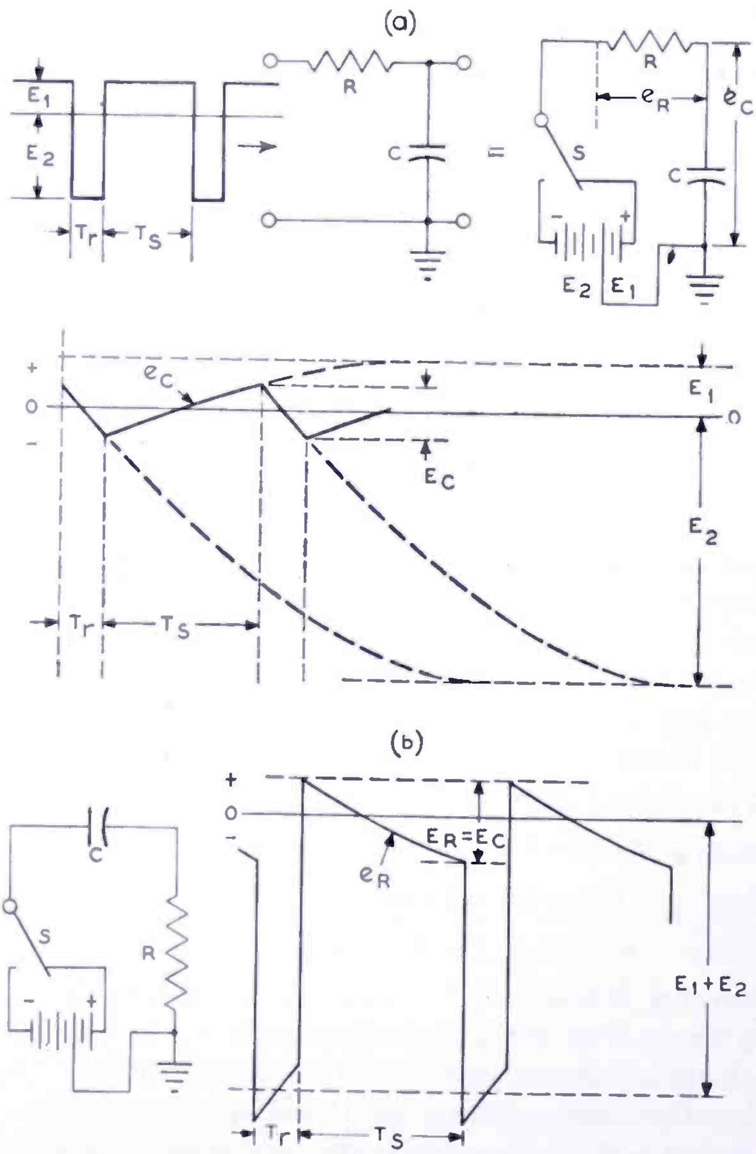


Fig. 16—Generation of control voltage for booster triode, V_2.

Efficiency and Power Feedback

The efficiency and power output of deflection circuits are relatively low. The overall efficiency can at best equal the square of the normal oscillator efficiencies since the a-c power output of tube V_1 must again be controlled and rectified by a second tube V_2. The booster tube V_2 is thus in principle a rectifier. An operating efficiency of 80 per cent per tube and a similar circuit efficiency for the deflection coils and transformer would give an overall efficiency of 50 per cent.

The efficiency of circuits in actual practice may be considerably lower than this value and, in fact, equals zero when the d-c power output, $E_{B_2} \times I_{b_2}$, is dissipated in a bypassed resistor (Figure 17), which replaces the battery $E_{B_2} = E' - E_2$ in Figure 8(b).

It is possible, however, to feed the rectified power back into the power source or power tube by use of circuits employing proper trans-

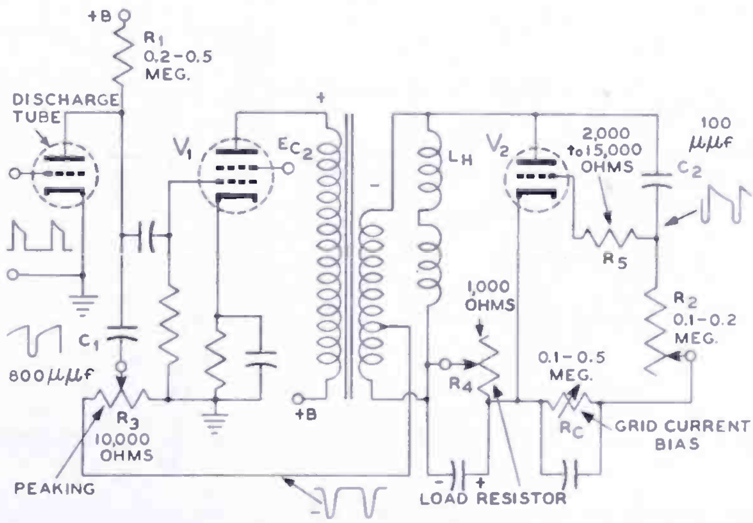


Fig. 17—Practical circuit for generating control voltage for booster triode, V_2.

former ratios. The use of such a circuit is justified, even for small power gains, if important operating advantages are secured. Various developments prior to and during the war indicated that as the definition of television systems is increased, the kinescope and pick-up tubes used require increasingly larger deflection energies. Such tubes would be impractical without some method of energy conservation. Based on investigations of power feedback, the author successfully demonstrated in 1944 a high-definition color television system utilizing series power feedback.

Parallel Power Feedback

Feedback of secondary d-c power into the power supply source requires the matching of voltages by means of a transformer with the proper step-up ratio. The transformer is given a step-up ratio $n_2/n_1 > 1$ which is adjusted to give $E_{B_2} = E_{B_1}$ with linear deflection current.

This ratio is: $$n_2/n_1 \simeq 1 + (E_1 + E_2)/E_L \tag{19}$$

The parallel connection is shown in Figure 18.

The 3/2-power relation of current and voltage in electron tubes causes a change in the voltage ratio $(E_1 + E_2)/E_L$ for changes in current amplitude requiring adjustment of the transformer ratio. This is also true for frequency changes. The parallel feedback circuit, therefore, is of little practical interest.

Series Power Feedback (Booster Circuit)

Power feedback in series with the power source requires matching

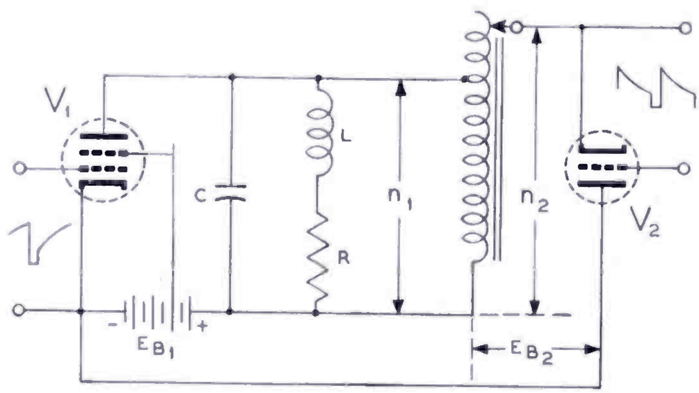


Fig. 18—Circuit with parallel connection for power feedback.

of average currents by adjustment of the transformer ratio. The transformer is given a step-down ratio, $n_1/n_2 < 1$, which is adjusted to give equal plate currents $I_{b_2} = I_{b_1}$ with linear operation. (Figure 19 (a)). After series connection of the secondary load terminals (i.e. the storage capacitance C_B) with the power source, the voltage E_{B_1} of the power source can be reduced to $E_{B_1}' = E_{B_1} - E_{B_2}$ (Figure 19(b)). The transformer ratio is: $$n_1/n_2 = I_{b_1}/I_{b_2}' \tag{20}$$

where I_{b_2}' is the current obtained with $n_1/n_2 = 1$.

The ratio is independent of current amplitude and frequency as the circuit Q and wave shapes are substantially constant. Stability of operation at all amplitudes is obtained by an essentially independent screen-grid voltage. The circuit with series power feedback is, therefore, a practical step toward the ideal wattless deflection circuit and, thus, deserves further discussion.

Adjustment of the transformer ratio is, in effect, a linearity control. A smaller step-down ratio giving $I_{b_2} < I_{b_1}$ with linearity before series connection will cause overdamped operation after series connection (Figure 19(c)) because I_{b_2} is forced to equal the pentode-controlled current I_{b_1} (equivalent to a decrease of R in Figure 19(a)).

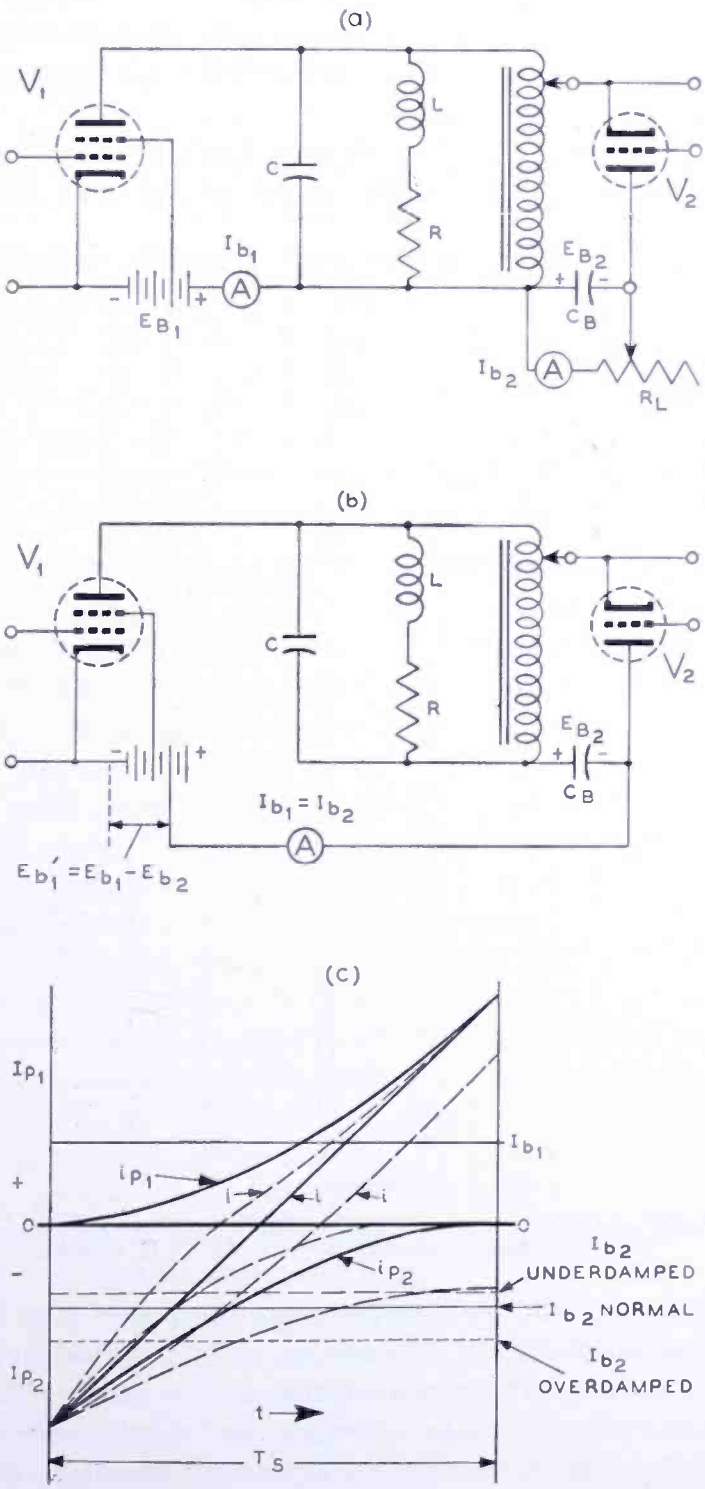


Fig. 19—Circuit with series connection for power feedback; and, effect of adjusting transformer ratio on linearity of current (i).

A larger step-down ratio causes $I_{b_2} > I_{b_1}$ with linearity before series connection, and after series connection, will give the underdamped condition indicated in Figure 19(c).

Small errors in linearity in a circuit with fixed transformer ratio can be corrected by adjustment of the grid-voltage amplitude or grid bias on V_2.

The transformer should have slightly less step-down for this purpose. As the current ratio is determined by the Q of the circuit, a

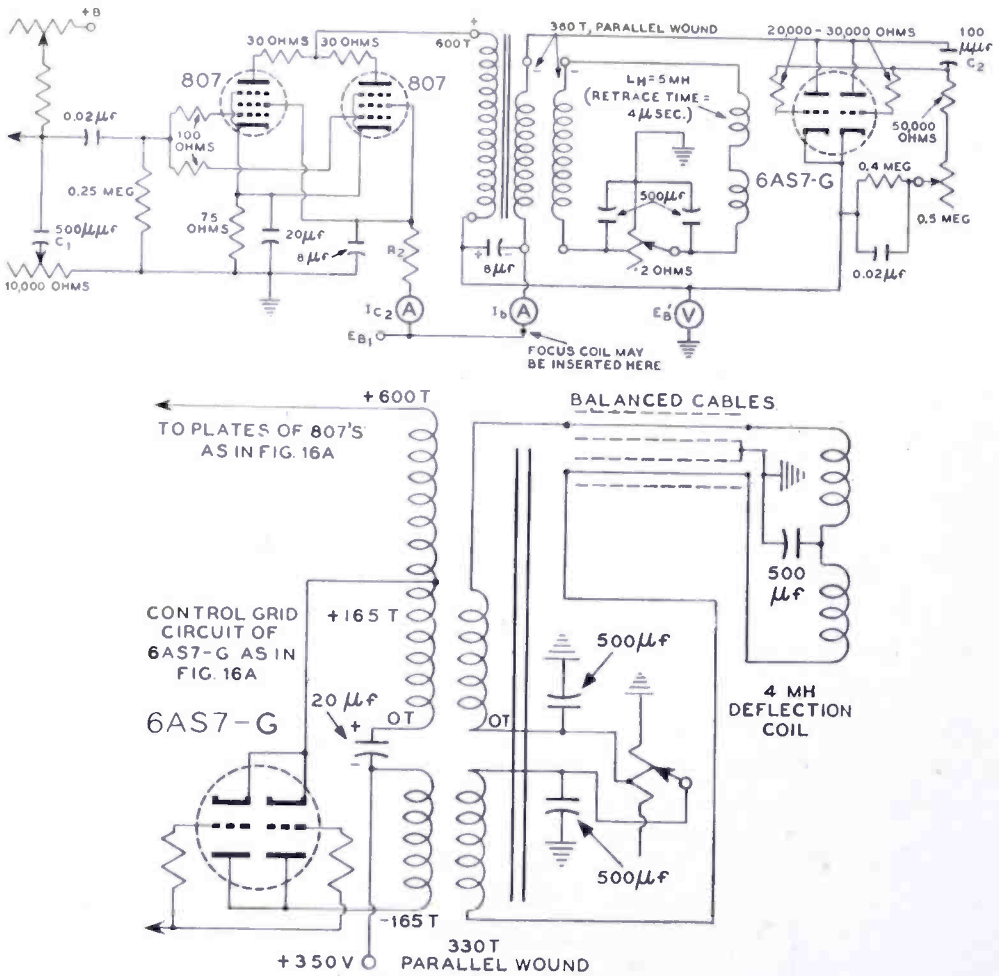


Fig. 20—Practical circuits with power feedback for balanced and unbalanced deflection-coil connections with 6AS7-G triode.

change in linearity will be observed when the power tube V_1 is made to load the circuit during T_r by reduction of its negative cutoff or peaking voltage. It must be kept in mind that only a change of the peak-to-peak to average current ratio will effect the linearity and when plate-current cutoff is maintained during T_r, bias or amplitude adjustments on V_1 will, therefore, be ineffective in changing the linearity of circuit operation.

Table I—Operation characteristics for the circuit of Figure 20(a).

Horizontal Deflection Frequency	15.75 KC		31.5 KC		Remarks
E_{B_1} (volts)	315	350	315	350	The maximum deflection amplitude at 15.75 kilocycles is approximately 30 per cent larger than at 31.5 kilocycles.
E_B' (volts)	350	382	490	520	
I_b (amperes)	0.175	0.195	0.135	0.14	
I_{c_2} (amperes)	0.025	0.022	0.029	0.026	
R_2 (ohms)	625	1650	625	1650	
Total Plate Power Input (watts)	61	75	66	73	$I_B \times E_B'$
Power Gain (watts) ...	6.1	6.2	23.6	23.8	$I_b\,(E_B' - E_{B_1})$
Plate Power Efficiency (per cent)	10	8.4	36	33	$(E_B' - E_{B_1})/E_B'$

This circuit property results in remarkable stability of deflection linearity. The circuit is further unique in being capable of linear operation at low and high frequencies; a large frequency change in circuits with booster triode requires only proportional changes of capacitance in the grid-voltage generating circuits (C_1 and C_2 in Figure 20(a)) to maintain normal grid-voltage wave shapes.

At low frequencies or small amplitudes, the booster rectifier tube V_2 consumes power from the supply E_{B_1} because the inductive voltage E_L is smaller than the tube drop E_2, thus causing E_{B_2} to be negative.

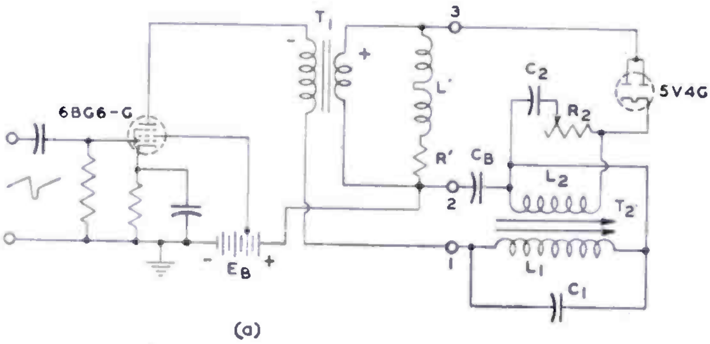
At high frequencies or large amplitudes, the inductive voltage becomes large; E_{B_2} is, therefore, positive, automatically meeting the requirement for increased plate supply voltage on V_1. This action protects the power tube V_1 against excessive plate dissipation when the grid signal is removed since the voltage E_{B_1} has a moderate value.

The step-down ratio of practical circuits varies between 1.4 and 1.7 according to the Q value. It is further increased in circuits where part of the boosted power is utilized for light external shunt loads, such as discharge circuits. The value of the inductive load has little effect (none for zero leakage reactance) on the transformer ratio and is determined by retrace-time considerations as in normal circuits. The voltage boost $E_{B_2} = E_L - E_2$, however, increases with load inductance, deflection frequency, and deflection current (see Equation (12)).

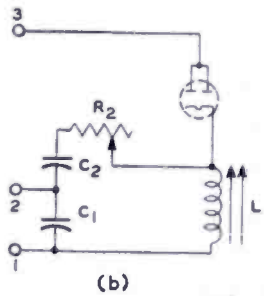
Examples of practical high power circuits for balanced and unbalanced deflection-coil connections are shown in Figures 20 and 21. Operation characteristics for the circuit of Figure 20(a) are given in Table I. Grounded deflection coils require impedance coupling or an isolated winding parallel with the V_2 winding (2 wires at the same time) to eliminate leakage reactance.

Excitation of a focus coil (for kinescopes) can be obtained by a series connection, as indicated in Figure 20(a). Adjustment of excitation by a shunt resistor will not disturb the operation of this circuit.

The efficiency of practical circuits depends on the availability of electron tubes with low plate-voltage loss (E_1 and E_2) and the construction of efficient transformers and deflection coils. The table in Figure 20 shows that efficiencies up to 33 per cent have been obtained



Circuit (a)
$L' = 5$ millihenries (deflection coil winding); $T_o = L_1$ and L_2 wound parallel (#29 Formex) series-connected as shown to obtain tight coupling with permeability tuning.
$L_1 = 2$ to 7 millihenries
Turns ratio of $L_o/L_2 = 1.6$
$L_o = 0.9$ to 2.7 millihenries
$C_2 = 0.05$ microfarads; $R_o = 500$ ohms variable
$C_1 = 0.015$ microfarads;
$C_B = 1$ to 4 microfarads.



Circuit (b) combines C_B and C_2, but is restricted by capacitive tap to higher Q-values and more critical operating conditions (Class B) than circuit (a). Ratio C_o/C_1 adjusts stepdown.

Fig. 21—Circuit with power feedback and controlled triode.

with 6AS7-G booster triodes and conventional transformer designs. The efficiency of the diode circuits (Figures 21(a) and (b)) is similar, as the power dissipation in R_2 is compensated by the somewhat lower power dissipation in the diode. The measured plate voltage boost $E_B' - E_{B_1}$ is higher; the effective boost, however, is reduced by the voltage drop across Z which varies from 1/2 peak-to-peak voltage $(ET/2)$ for $R_2 = 0$ to nearly peak-to-peak voltage E for large R_2 values.

Switch Circuits with Positive Resistance

The generation of linear sawtooth currents requires the closed-switch circuit to have the characteristic of a negative resistance $-R$. It has been shown that circuits with a diode (Figure 7) can function as $-R$ by inefficient operation with a large circulating current, or with good efficiency by plate voltage control with auxiliary transformer (Figure 14). Switch circuits with decreasing diode current *and* plate voltage (V_2) or with increasing plate voltage on V_1 have a positive resistance. Linearity is, therefore, not obtainable with many circuits although good current efficiency (ηi) is possible.

One type of circuit utilizes the grid and cathode of a triode power tube V_1 as a diode (V_2) by means of transformer coupling in a self-excited circuit.[3] Plate and grid currents are controllable by varying the space potential in the control-grid plane. The desired change in current distribution, i.e., a rising plate current and decreasing grid current, however, requires a rising plate voltage and, hence, a positive resistance characteristic of the switch circuit. Since the inductive voltage E_L must decrease during T_s, current linearity is not obtainable in the particular circuit without auxiliary potentials.

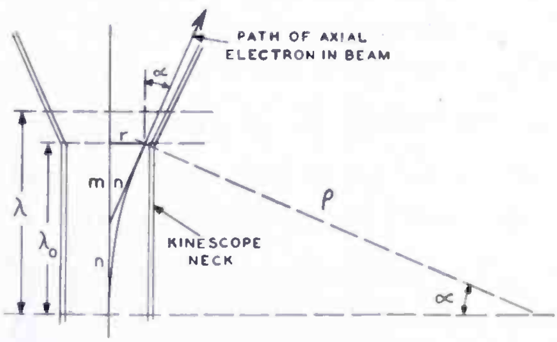


Fig. 22—Path of electron beam through transverse deflection-coil field of kinescope. ($m = n \cos \alpha$; $n = r/\sin \alpha$; $\lambda_0 = m + n = r\,(1 + \cos \alpha)/\sin \alpha$).

Because of difficulties in maintaining stable grid characteristics, it is, in general, undesirable to operate small electron tubes having oxide-coated cathodes with heavy grid currents.

Design Constants for Kinescope Deflection Systems with Matching Transformers

Transformer ratios other than those given for the circuit with power feedback may be desirable because of preferred tube or circuit operating conditions. The evaluation of numerical constants is necessary to indicate the range of currents and voltages in practical circuits and to indicate desirable constants for the electron tubes.

---

[3] Sawtooth current oscillations (non-linear)—
(a) T. H. Mulert and H. Baehring; Transformator-Kippgeraete Hansmitteilungen der Fernseh AG; Vol. 1, No. 3, pp. 82-88, April, 1939.
(b) L. R. Malling; Triode linear sawtooth-current oscillator *Proc. I.R.E.*, Vol. 32, No. 12, December, 1944.

Deflection coil constants

The electron beam in a kinescope passes through the transverse deflection-coil field of intensity H (see Figure 22) with the volt velocity E (anode potential).

$$H = 0.4\,\pi\,NI/l_o \tag{21}$$

In a magnetic field of constant intensity, the electrons follow a circular path with the radius

$$\rho = 3.33\,\sqrt{E}/H \tag{22}$$

The transverse field must be limited to a depth λ_o to obtain an electron path leaving the field at a distance r from the axis at an angle α. The field depth λ_o below the neck junction is limited to

$$\lambda_o = r\,(1 + \cos\alpha)/\sin\alpha \tag{23}$$

The radius r (Figure 22) must provide clearance for beam thickness and glass neck tolerances.

With $\rho = \lambda/\sin\alpha$ from Figure 22, Equations (21) and (22) yield an expression for the ampere turns required for beam deflection over the angle as follows:

$$NI_{(\alpha)} = 2.65\,l_o \sin\alpha\,\sqrt{E}/\lambda \tag{24}$$

The inductance of this winding is

$$L = 4\pi\,N^2 A/l_o\,10^9 \text{ (Henry)} \tag{25}$$

Equivalent dimensions for use of Equations (24) and (25) with practical coil designs are shown in Figure 23 where

l_o = air-gap length or inside diameter of iron shell in centimeters
λ = equivalent coil length in centimeters
$A = d \times \lambda$ = average cross section of flux in square centimeters

The coil winding is given approximately a constant number of turns per unit of projected length (l'). Modifications in field strength over the neck center or at the coil ends may be required for obtaining a rectangular trace on the kinescope screen depending on its radius of curvature and on the coil leakage field. (A large screen radius requires an increased effective length $\lambda \cong 1.2\,\lambda_o$ for 55-degree angle kinescopes

The equivalent coil field length changes inappreciably when the coil ends are folded up (Figure 23). Folding at the front end, however, permits an increase of winding length and pole-face length in the direction of the kinescope screen for a given value λ_o, and results in an increased effective length $\lambda \simeq 1.2\,\lambda_o$ for 55-degree angle kinescopes ($\alpha = 27.5$ degrees). The following calculations for a 55° deflection angle were made before the current 50° deflection angle was adopted, but they serve equally well to illustrate the method.

Folding of the back end does not permit a further change of length; it only shortens the coil physically. The leakage-field shape obtained with folded coil fronts reduces defocusing of the beam at large angles α and is, therefore, highly desirable.

When the design constants for wide-angle kinescopes are $\alpha = 27.5$ degrees, $r = 1.27$ centimeters, Equation (23) furnishes $\lambda_o = 5.18$ centimeters and with folded coils $\lambda = 6.2$ centimeters. The practical coil assembly (Figure 23) furnishes $l_o = 6$ centimeters and employing Equation (24) the required ampere-turns for the half-angle are determined as follows:

$$NI_{(27.5°)} = 1.18\,E \qquad (24a)$$

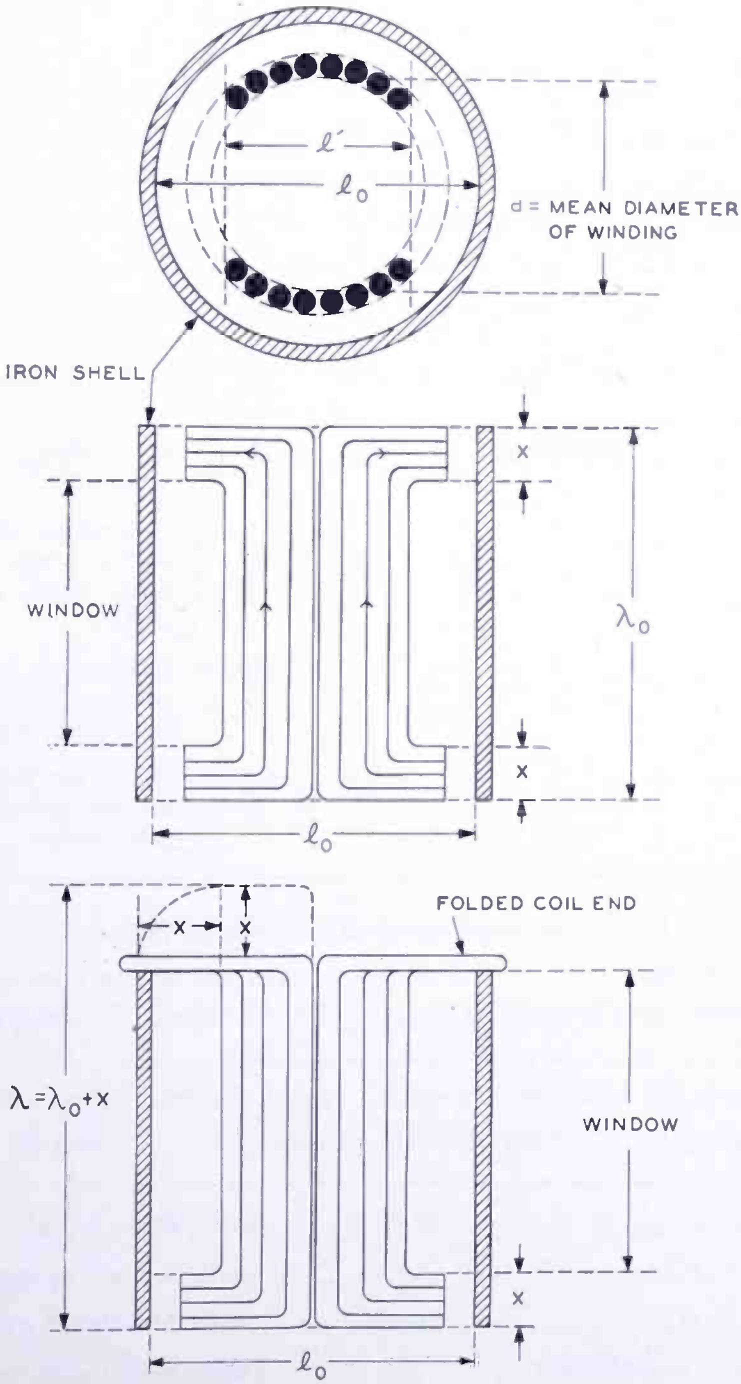


Fig. 23—Cross-sectional views of practical coil for magnetic deflection.

The value obtained by Equation (24a) must be multiplied by two for the peak-to-peak deflection over the full angle 2α.

The line length on the kinescope screen is approximately 83 per cent of the nominal tube diameter for present curved-face tubes. A picture size with a diagonal equal to the nominal tube diameter D, requires the deflection length $H = 0.8D$ and $V = 0.6D$ for a 4:3 aspect ratio. This furnishes

$$NI_{(H = 0.8D)} \simeq 2.28\, E^{\frac{1}{2}} \tag{24b}$$

and similarly $NI_{(V = 0.6D)} \simeq 1.95\, E^{\frac{1}{2}}$ (λ_o is shorter for vertical coils.)

For obtaining good Q values at the retrace frequency $f_o = 1/2\, T_r$, the reduction of eddy-current losses in copper and iron requires small wire diameters in both windings and the iron shell as shown by Table II.

Table II—Q-Measurements on Horizontal Deflection Winding at 87 kilocycles.

Test No.	Wire Size	L(mh)	Q	Condition of Test
1	#29 SSE	4.42	50	H-winding by itself, no iron.
2	"	5.2	8	H-winding inserted in heavy wire vertical coils with iron wire shell.
3	"	5.2	10.7	H-winding in fine wire vertical coil (#29) with iron wire shell.
4	"	4.8	5.8	Same as (3) but connected to transformer secondary.

Transformer Ratio and Electron-Tube Characteristics

The capacitance C_o of the deflection system is the sum of the circuit element capacitances or their reflected values. It determines the

Table III—Estimated value of circuit element capacitances.

Circuit Element	C (micromicrofarads)	Remarks
Deflection coil	60-80	with short leads
Plate of V_1 (C_p)	10	with top cap connector (807)
Plate of V_2 (C_p)	20-40	with associated grid circuit
V_2 filament transformer....	25	used only for direct coupling
Transformer or choke	20-50	depending on winding method

Circuit constants for $Q = 6$, E(kinescope) $= 10$ kilovolts, $NI = 228$ ampere turns, and transformer constants as shown in Figure 24.

	$N_T = n_1/n_2$	C_o $(\mu\mu f)$	L_o (mh)	L_H (mh)	N_H	$I = \hat{\imath}_1 + \hat{\imath}_2$ (amp)	$\hat{\imath}_2$ (amp)	I_{b_2} (amp)	E_{L2} $(volt)$	$\hat{e}_2$ (kv)	$\hat{\imath}_1'$ (amp)	I_{b_1} (amp)	E_{L1} $(volt)$	$\hat{e}_1$ (kv)	807 E_{bb} min $(volts)$
Low C_o	1.	135	24.4	24.4	700	0.326	0.16	0.063	137	2.3	0.20	0.090	137	2.3	240
	1.5 ...	70	47.	18.8	615	0.370	0.18	0.072	120	2.	0.154	0.070	190	3.2	290
	2.	50	66.	14.8	546	0.417	0.20	0.080	106	1.8	0.130	0.060	224	3.8	320
	2.5 ...	43	76.5	11.	470	0.485	0.225	0.092	92	1.6	0.120	0.054	243	4.1	340
	3.	39	84.5	9.4	435	0.525	0.25	0.100	85	1.5	0.110	0.050	270	4.6	370
High C_o	1.	200	16.5	16.5	576	0.395	0.19	0.076	112	1.9	0.245	0.110	112	1.9	...
	1.5 ...	115	28.5	11.4	480	0.475	0.23	0.092	94	1.6	0.197	0.090	150	2.5	250
	2.	90	36.5	8.	402	0.567	0.27	0.008	78	1.3	0.175	0.079	165	2.8	265
	2.5 ...	79	41.5	5.9	345	0.662	0.32	0.128	68	1.2	0.165	0.074	178	3.	275
	3.	73	45.	4.45	300	0.760	0.36	0.144	58	1.	0.157	0.071	185	3.2	285
		retrace constants		deflection coil constants			V_2 load constants				V_1 load constants				

Table IV—The characteristic requirements for circuit elements in power tubes (V_1 and V_2) as a function of transformer ratio.

permissible inductance L_0 of the system for a given retrace time (Equation 1). The circuit capacitances vary considerably with manufacturing and assembly methods. Estimated values are given in Table III.

The resultant effective capacitance C_0 for the two extremes of capacitance in Table III and the inductance L_0 have been computed for various transformer ratios between V_1 and V_2 for a normal television retrace time $T_r \simeq 6$ microseconds and scansion $T_s \simeq 58$ microseconds. Maximum and minimum values of C_0 are given in Table IV.

The deflection coil current $(\hat{\imath}_1 + \hat{\imath}_2)$ is computed for $NI = 228$ as given by Equation (24b) for $E = 10$ kilovolts. The load constants for the tubes V_1 and V_2 are computed for the transformer network given in Figure 24, and include, therefore, a 10 per cent exciting current

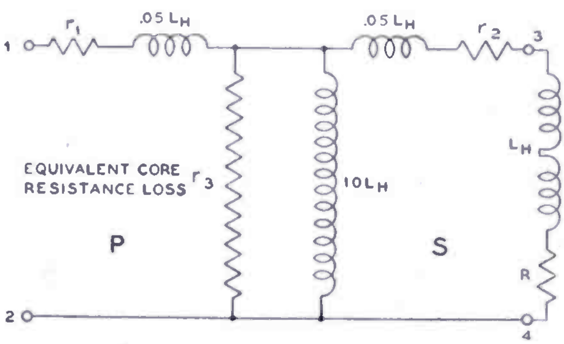


Fig. 24—Transformer network with representative load constants for V_1 and V_2.

and the effects of a 5 per cent leakage reactance per winding. The average plate currents are approximately $I_{b_1} \simeq 0.45\,\hat{\imath}_1$ and $I_{b_2} \simeq 0.4\,\hat{\imath}_2$.

Power Tube Characteristics and Supply Voltage

Table IV establishes the characteristic requirements for circuit elements and power tubes (V_1 and V_2) as a function of transformer ratio. Pulse voltage $(\hat{e}_2)$, plate current $(\hat{\imath}_1')$, and effects of yoke capacitance decrease with increasing step-down ratio. The pulse voltage $\hat{e}_1$, however, increases and in practical circuits it is often 25 per cent higher than given because of sectional resonances. The power tube requires, therefore, a top cap plate connection. Practical designs without power feedback may correspond to the higher capacitance values (2nd half of Table IV) which require higher currents and lower voltages. Designs with power feedback require a small step-down ratio to the tube V_2. The supply voltage E_{bb} min in the last column

Table V—Characteristics of low-mu twin power triode 6AS7-G.

Electrical Characteristics:

Heater, for unipotential Cathode:

Voltage (a-c or d-c)	6.3	....	Volts
Current	2.5	....	Amperes

Mechanical Characteristics:

Mounting Position	Any
Maximum Overall Length	5-5/16 inches
Maximum Seated Length	4¾ inches
Maximum Diameter	2-1/16 inches
Bulb	ST-16
Base	Medium Shell Octal 8-Pin

D-C Amplifier

Values are for each unit

Maximum Ratings, Design-Center Values:

Plate Voltage	250 max.	....	Volts
Plate Current	125 max.	....	Milliamps
Plate Dissipation	13 max.	....	Watts
Peak Heater-Cathode Voltage:			
Heater negative with respect to cathode	300 max.	....	Volts
Heater positive with respect to cathode	300 max.	....	Volts

Characteristics:

Plate-Supply Voltage	135	....	Volts
Cathode-Bias Resistor	250	....	Ohms
Amplification Factor	2.0	....	
Plate Resistance	280	....	Ohms
Transconductance	7000	....	Micromhos
Plate Current	125	....	Milliamps

Maximum Circuit Values (for maximum rated conditions):

Grid-Circuit Resistance:

For Cathode-Bias operation†	1.0 max.	....	Megohm

Booster Scanning Service

Values are for each unit

Maximum Ratings, Design-Center Values:

Peak Inverse Plate Voltage*	1700 max.	....	Volts
Plate Current	125 max.	....	Milliamps
Plate Dissipation	13 max.	....	Watts
Peak Heater-Cathode Voltage:			
Heater negative with respect to cathode	300 max.	....	Volts
Heater positive with respect to cathode	300 max.	....	Volts

Maximum Circuit Values (for maximum rated conditions):

Grid-Circuit Resistance:

For Cathode-Bias operation†	1.0 max.	....	Megohm

* The duty cycle of the peak inverse voltage pulse must not exceed 15% of one scanning cycle and its duration must be limited to 10 microseconds.

† Operation with fixed bias is not recommended.

is approximated for normal transformer and coil resistances by

$$E_{bb} \text{ min} \simeq E_1 + 1.15\, E_{L_1} \tag{26}$$

E_1 is the peak-current tube-drop in an 807 power tube. Experience has shown that further allowances in voltage must be made for blanking margins, for variation in tubes to reduce screen-grid dissipation, and for a self-bias voltage which raise E_{bb} min by approximately 20 per cent.

Booster Triode Characteristics

To meet the characteristics and requirements outlined here for a booster tube, the 6AS7-G was developed. This tube was designed by the author and produced during the war for use in military television and radar equipment. It is now available and has been successfully utilized in the type of circuits suggested for magnetic deflection in this paper. Its characteristics are given in Table V (see preceding page.)

EFFECT OF FIELD STRENGTH ON DIELECTRIC PROPERTIES OF BARIUM STRONTIUM TITANATE*

BY

H. L. DONLEY

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—*Results are given on the dependence of dielectric properties on field strength for the following titanate compositions:*

% $BaTiO_3$	% $SrTiO_3$
69	31
80	20
90	10

over the temperature range —40 to 80 degrees centigrade and frequency range extending in general from direct-current to 1 megacycle.

The high-dielectric-constant ceramics, analogous to ferromagnetic materials, show saturation effects of dielectric flux density with field strength as well as a critical or Curie temperature above which the non-linearity of dielectric constant with field strength disappears.

The pronounced non-linear characteristics observed for these polycrystalline materials suggests their possible practical use as non-linear circuit elements. Some results are given on the use of these titanate capacitors as the non-linear element in a frequency multiplier, frequency changer, and frequency modulator.

INTRODUCTION

IN the usual dielectric material the dielectric constant is independent of the applied voltage except at extremely high field strengths. However, in the case of the high dielectric constant materials, particularly barium titanate, a very strong dependency of dielectric constant upon field strength has been observed.[1] In fact barium titanate exhibits hysteresis and saturation effects[2] analogous to the corresponding properties of ferromagnetic materials. Also, like ferromagnetic materials, barium titanate displays maximum (initial or low

---

* Decimal Classification: R216.1 × R381.12.

[1] See, for example, B. M. Wul and I. M. Goldman, "Dielectric Constant of Barium Titanate as a Function of Strength of an Alternating Field", *Compt. Rend. Acad. Sci.*, Vol. 49, pp. 177-180, October, 1945.

[2] A. de Bretteville, "Oscillograph Study of Dielectric Properties of Barium Titanate", *Jour. Amer. Ceramic Soc.*, Vol. 29, pp. 303-307, November, 1946.

field) dielectric constant at a critical or Curie temperature above which the dielectric constant rapidly decreases. The non-linearity between dielectric constant and field strength and hysteresis phenomena are evident only at temperatures below the critical or Curie temperature.

Somewhat above the Curie temperature the dielectric irregularities disappear, and the dielectric shows properties similar to those of the familiar insulator. In addition to barium titanate various compositions of barium strontium titanate also exhibit similar non-linear dielectric properties with field strength. The accumulated evidence[3,4,5] points to the proper classification of these materials as ferroelectrics, where the term ferroelectric implies the anology of the dielectric properties to the magnetic properties of ferromagnetic materials. The classification of these materials however, is based upon measurements on polycrystalline material only and consequently the complete story awaits the production of single crystals of these materials.

From the standpoint of possible practical application, the non-linear properties of the barium strontium titanate solid solution series may assume considerable importance for the following reasons: (a) The Curie temperature may be readily shifted simply by varying the composition of the titanate; the range in Curie temperatures extends from about 130 degrees Centigrade for barium titanate to around −200 degrees Centigrade for strontium titanate; this means then that the non-linear or ferroelectric properties may be varied through rather wide ranges;* (b) In any practical application, the marked degree of non-linearity of dielectric constant with applied field manifested by these polycrystalline ceramics eliminates the necessity of specifying a definite crystallographic direction for the observance of the non-linear effect, as is the case, for instance, in Rochelle salt — the outstanding example of a ferroelectric material; (c) The dielectric losses accompanying the non-linearity of dielectric constant with field are reasonably low unless the material is driven to the approach of saturation; (d) The extremely high dielectric constants (on the order of 5000 — 10,000) of these materials makes possible physically small

---

[3] A. von Hippel, R. G. Breckenridge, F. G. Chesley and L. Tisza, "High Dielectric Constant Ceramics", *Ind. Eng. Chem.*, Vol. 38, pp. 1097-1109, November, 1946.

[4] B. M. Wul, "High Dielectric Constant Materials", *Jour. Phys.* (U.R. S.S.), Vol. 10, No. 2, pp. 95-106, 1946.

[5] V. Ginsburg, "On the Dielectric Properties of Ferroelectric Crystals and Barium Titanate", *Jour. Phys.* (U.S.S.R.), Vol. 10, No. 2, pp. 107-115, 1946.

* The ferroelectric properties apparently disappear for strontium titanate alone; see Reference 3.

elements which fact assumes practical importance when physical size must be considered; (e) The persistence of the non-linear properties of these materials to high radio frequencies is also of practical importance; and (f) The ease of manufacture of these materials by a simple firing process† emphasizes the interest in these materials in spite of their rather variable properties.

The present report is limited primarily to the presentation of data of some engineering interest on the observed non-linearity between dielectric properties and field for the titanate compositions 69% $BaTiO_3$ — 31% $SrTiO_3$ (69/31); 80% $BaTiO_3$ — 20% $SrTiO_3$ (80/20); 90% $BaTiO_3$ — 10% $SrTiO_3$ (90/10) over the temperature range −40 to 80 degrees Centigrade and frequency range extending in general from direct current to 1 megacycle. It must be emphasized that the characteristics and data presented are qualitative in the sense that the results apply to polycrystalline materials only. Consequently, the results given may be regarded as overall or average values for the compositions studied. In addition, no account has been taken of the influence of possible mechanical resonance phenomena on the observed non-linear dielectric characteristics except in a general way as evidenced by widely separated test frequencies.

This paper serves also to give some results of the specific application of these materials as non-linear circuit elements.

DEFINITIONS — METHODS OF MEASUREMENT

The fact that the dielectric constant of the materials under study depends upon the field strength necessitates a careful definition of dielectric constant. Since the anology between the dielectric behavior of these titanates and ferromagnetism is a fairly complete one, it is natural to choose definitions for the dielectric constant similar to those for magnetic permeability. Following Cady[6] the normal or overall dielectric constant, k, is defined as the ratio of induction D to the maximum applied field strength E where D varies over the hysteresis cycle. The differential dielectric constant, k_d — the slope of the hysteresis curve at any point, becomes the initial dielectric constant, k_o, for $E = O$ or E small. Where a small a-c field is applied at

† The general features of the firing process and preparation of samples are adequately described elsewhere and will not be of concern here; see Reference 3; also E. N. Bunting, G. R. Shelton and A. S. Creamer, "Properties of Barium Strontium Titanate Dielectrics", *Jour. Research* (National Bureau of Standards), Vol. 38, No. 3, pp. 337-349, March, 1947.

[6] For a more complete discussion of this analogy than is given here, see W. G. Cady, PIEZOELECTRICITY, pp. 550-551), McGraw Hill Book Co., New York, N. Y., 1946.

any point on the $D:E$ diagram the small variations in D show no hysteresis, and one obtains the reversible dielectric constant, k_r.

The value of k_r can be readily obtained by a-c bridge measurements but such measurements become unreliable when k is desired for large a-c fields because of the distortion effects produced by the sample under measurement. The oscillograph provides the best experimental means for obtaining the various dielectric constants, but quantitative determinations are difficult particularly at the higher frequencies. The normal k or a-c dielectric constant, the only dielectric constant except in connection with Figure 6 measured in the following, was obtained by parallel tuned circuit measurements. In particular for measurements on thick samples a step-up transformer was used in which the sample under investigation and a calibrated variable condenser tuned the high-voltage secondary, the low impedance primary being driven by an 807 tube operated Class A. Circuit resonance and voltage drop across the unknown condenser were indicated by means of an auxiliary sharply-tuned resonant circuit loosely coupled to the transformer secondary.

Variation of Curie Temperature with Composition and Dielectric Properties as a Function of Temperature

In a general way the course of the dielectric constant and losses with temperature for the various compositions of the titanate dielectric is as illustrated in Figure 1 for the observed results for the particular composition 69/31. The very rapid change in dielectric properties near the critical or Curie temperature is characteristic of these materials. As already mentioned, the position of the dielectric constant peak on the temperature scale is determined primarily by the titanate composition. The exact position and magnitude of the peak in dielectric constant is also a function of firing temperature, degree of porosity of the fired body, field strength, and other factors which are not considered here.

The Curie temperature as a function of the barium strontium titanate composition is shown in Figure 2. It is seen that as the strontium content of the mixture is increased the position of k_{max} moves down in temperature in approximately a linear manner.

Effect of Field Strength on the Composition 69/31

The 69/31 mixture was chosen for the most complete study because the Curie point conveniently comes at just about room temperature,

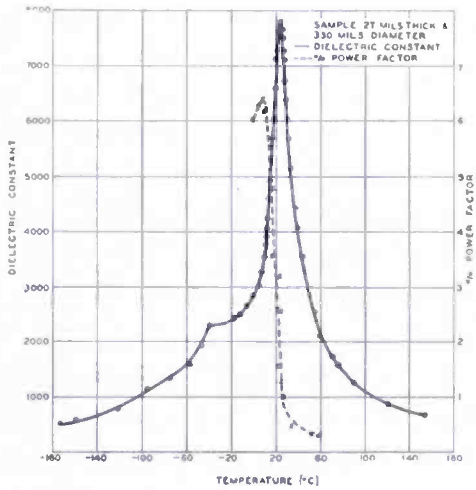


Fig. 1—Dielectric properties of the 69/31 mixtures at 1 kilocycle and 120 volts/centimeter.

at which point a dielectric constant of about 9000 is developed at 1 kilocycle as shown in Figure 1.

By means of a ballistic galvanometer, the effect of a d-c field on a 30-mil thick sample was observed for successively increasing voltages in which a 5-second charging time was employed, followed by a 5-minute short circuit of the sample before the next voltage application. The results obtained are shown in Figure 3, in which the evidence of hysteresis is readily apparent from the difference in k values for increasing and decreasing voltage values. The corresponding flux density vs. E curves are also given in Figure 3, which shows that the material is far from saturated at 3 kilovolts/centimeter. This lack of saturation for this particular material has also been observed at much higher field strengths. A comparison of Figures 1 and 3 shows roughly

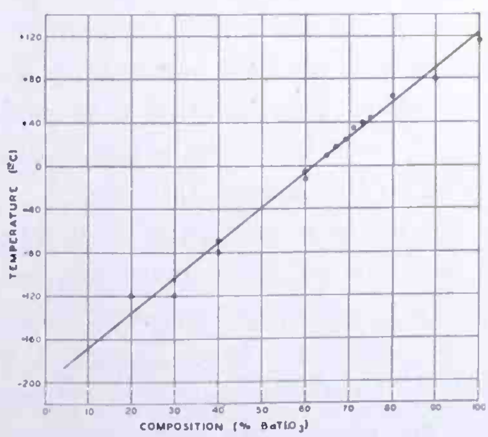


Fig. 2—Curie temperature as a function of barium strontium titanate composition.

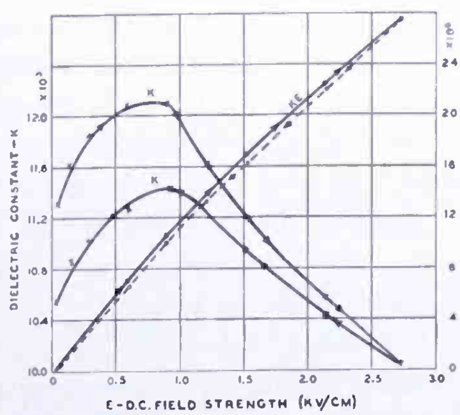


Fig. 3—D-C dielectric constant and flux density as a function of d-c field for the 69/31 mixture at room temperature.

the rapid decrease in initial dielectric constant in passing from d-c to a-c values.

Low-capacity samples, being the most convenient for high frequency measurements, consisted of irregularly-shaped pieces mounted by means of soldered leads connecting the fired-on silver electrodes and the base pins of a polystyrene cup which was filled with wax for moisture proofing the sample. The 27-mil thick low-capacity pieces were taken from the 330-mil diameter sample whose dielectric constant variation with temperature is given in Figure 1. The initial value of dielectric constant k_o for the small irregularly-shaped samples at a particular temperature was taken from the characteristic curve, Figure 1. The drop in k_o with frequency was readily determined from bridge measurements.

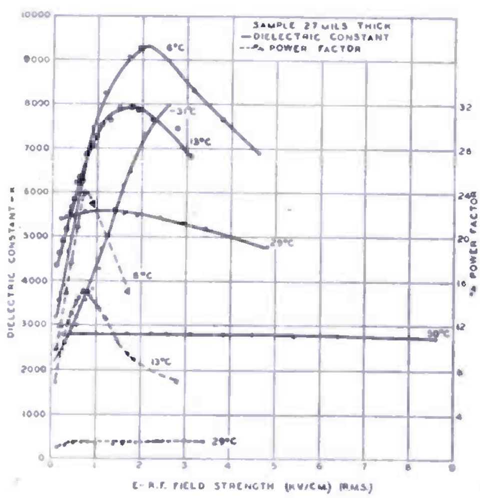


Fig. 4—Dielectric properties of the 69/31 mixture as a function of temperature and field strength at 10 kilocycles.

In order to obtain high a-c field strengths for only a few applied volts, samples about 1.5 mils thick were also studied particularly for the use in the applications mentioned later. Samples this thin could not be fired directly but small samples were taken from a larger disk which had been previously ground down to the desired thickness.

Some of the observed results for the 69/31 composition at 10 kilocycles, which illustrate the general behavior of these materials particularly in the ferroelectric region as a function of temperature and field strength, are shown in Figure 4. Although from Figure 4 an optimum maximum k at around 6 degrees Centigrade appears evident, the exact magnitude of the peak k is not too certain since the losses at k_{max} become high, making the determination of maximum capacity difficult. There are, however, some characteristics to be observed from

Figure 4 for the 69/31 mixture which apply generally to other compositions studied. These general features are: (1) with a decrease in temperature k_{max} moves to higher field strengths; (2) a more rapid rise in losses occurs with E at lower temperatures as well as a more rapid increase in k as the field strength is increased; (3) the position of k_{max} is decidedly voltage sensitive in addition to being temperature sensitive; (4) the peak of the loss characteristic with E occurs at lower field strengths than the peak of k with E; and (5) the variation of k with E disappears at temperatures somewhat above the critical or Curie temperature as seen from the 50-degree Centigrade curve in Figure 4.

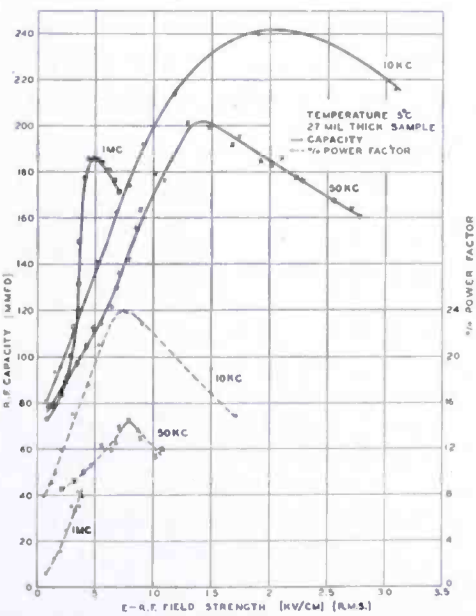


Fig. 5—Observed variation of capacity and power factor with field strength for the same sample of the 69/31 mixture showing shift of k_{max} to lower E as frequency is increased.

The general behavior, as brought out above, agrees with the observation of others[2,3,4] particularly on $BaTiO_3$ alone. Oscillograms of these materials show typical hysteresis loops which decrease in loop area as the Curie temperature is approached.

Observations of the non-linear dielectric properties with field were also taken at various temperatures at frequencies of 50 kilocycles and 1 megacycle for the same sample of the 69/31 mixture. Other samples of the same thickness revealed about the same behavior. Some typical characteristics for 5 degrees Centigrade at three different frequencies are given in Figure 5.

The decrease of k_{max} with E as the frequency is increased is a general characteristic also observed for other mixtures, as shown, for example, in Figure 8. The frequency at which the dielectric anomoly under consideration disappears entirely for these materials is not known at present. The non-linearity has been observed at 10 megacycles with only a few per cent drop in the k_{max} observed at 1 megacycle.

The gradual decrease in initial dielectric constant with frequency is roughly as shown in Figure 5. The drop in initial k is in general more rapid for audio than for radio frequencies.

The observed shift of k_{max} to lower E when the frequency is raised (Figure 5) may or may not be characteristic of the material itself, but merely evidence of the approach to mechanical resonance for the particular thickness of samples chosen. However, the same general feature observed for thin (1.5-mil) as for thick (27-mil) samples of this mixture as well as others throws some doubt on the presence of

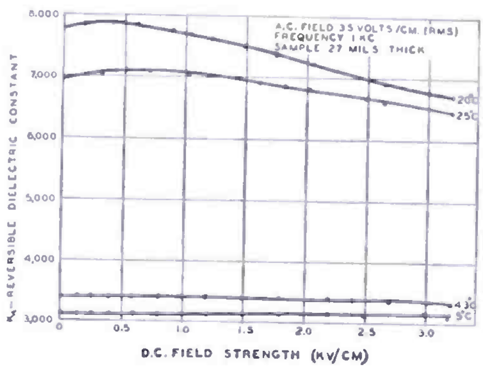


Fig. 6—Dependence of reversible dielectric constant on d-c field for the 69/31 mixture.

mechanical resonance (see Figure 8). Also the lower peak loss at higher frequencies casts additional doubt on the occurrence of mechanical resonance. On the other hand, in view of the fact that k_{max} shifts to lower E for d-c than for a-c (compare Figures 3 and 4), one might have expected that k_{max} should shift to higher E as the frequency is raised.

As pointed out previously, the reversible dielectric constant may be readily measured by means of an a-c bridge. Figure 6 shows the results of such measurements on the same 27-mil thick sample whose k–E characteristics are given in Figure 4. Figure 6 shows that the effect of d-c field is to lower the normal dielectric constant. Furthermore, the effect of the d-c bias is a maximum when k is a maximum, namely at the Curie temperature, and rapidly decreases in effect at temperatures above and below the Curie temperature when the dielectric constant is considerably reduced over the peak value.

Effect of Field on Compositions 80/20 and 90/10

As judged from Figure 2 the Curie temperature moves from about room temperature for the 70/30 mixture to about 60–70 degrees centigrade for the 80/20 mixture and roughly 80–90 degrees Centigrade for the 90/10 mixture. The dielectric constant temperature characteristics for these latter compositions have not been shown here, but if one translates the characteristic curve for 69 per cent $BaTiO_3$, Figure 1, to e.g. 60 degrees Centigrade for the peak k, then obviously a much lower temperature capacity coefficient results for room temperature for the 80 and 90 per cent $BaTiO_3$ compositions with dielectric

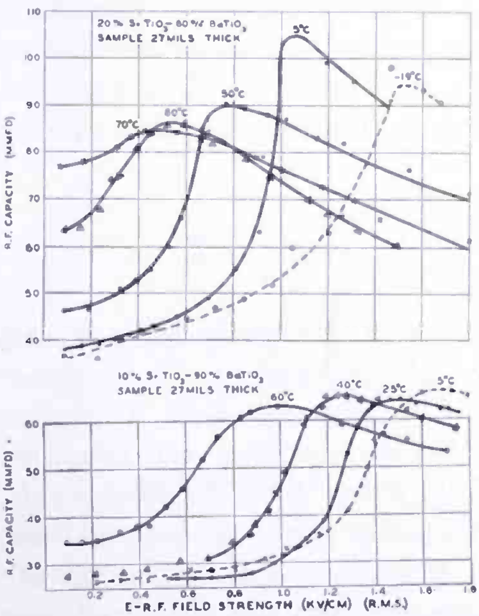


Fig. 7 — Radio-frequency capacitance as a function of field strength at 1 megacycle for mixtures 80/20 and 90/10.

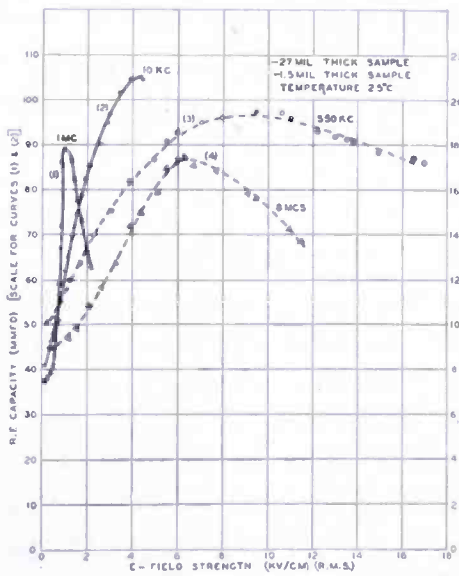


Fig. 8 — Dependence of capacity upon field strength at various frequencies for thin and thick samples of 80/20.

constants roughly on the order 2000–3000 at room temperature. Also, as previously noted, the degree of non-linearity of k with E increases at room temperature as the Curie point moves higher in temperature for the higher barium content mixtures.

The observed change of capacity as a function of field strength at 1 megacycle for various temperatures for the 80/20 and 90/10 compositions is given in Figure 7. The same general features as noted for the 69/31 mixture also hold for these compositions except a translation of the characteristics to higher temperatures. At a given temperature, the shift of k_{max} to higher E as the composition is changed toward $BaTiO_3$ alone is also brought out by comparison of the curves given in Figure 7 for the two compositions.

The shift of k_{max} to lower E as the frequency is raised was also observed for these latter compositions as for the 69/31 composition as shown by curves (1) and (2) of Figure 8. In addition, this same shift to lower E with increase in frequency was observed for thin (1.5-mil) samples (see Figure 8) as for the 27-mil thick samples.

The observed translation of k_{max} to higher E for thin as compared to thick samples is not surprising when one realizes that the possible presence of any minute low-dielectric-constant surface layers or gaps between the electrodes and the ceramic surface will have a more pronounced effect on the thin than on the thick samples. This is, of course, particularly true of these extremely high-dielectric-constant materials. Also, the presence of voids throughout the fired body presumably have a proportionately greater influence on the thin samples than the thick ones.

Applications

(1) *Frequency Multiplier*

One application of the non-linearity of dielectric constant with field strength of these materials is for the purpose of frequency multiplication. Such a non-linear dielectric possesses a distinct advantage over a vacuum tube non-linear resistance for frequency multiplication, because the conversion of power from the fundamental frequency to power at the harmonic frequency can occur with no losses if the dielectric is a pure reactance. Practically then, to realize this zero power loss feature means that the required low dielectric losses also do not change even when the material is driven to saturation. Obviously these materials far from fulfill this requirement, as a glance at their characteristic curves reveals. An idea of the extent to which these dielectrics do fulfill the above-mentioned advantage may be gained from the results of the following rather qualitative experiments.

As in the magnetic case, where sine-wave excitation is used, oscillograms of the resulting wave form arising from such a non-linear titanate condenser show essentially odd harmonics, of which the third and fifth are the most important. The use of a bias on the non-linear condenser results, of course, in even harmonic output the amplitude of which will depend upon the values of bias and radio-frequency amplitude.

The results given here are confined to third harmonic operation only, being those results, perhaps, of most practical interest.

Perhaps the most direct or practical circuit arrangement utilizing the non-linear capacitor for the present purpose is that shown in Figure 9 where the fundamental voltage, v_f across the parallel tuned circuit L_1C_1 is applied to $C(v)$, the non-linear capacitor. The high impedance to the harmonic is provided by the tuned circuit LC, which gives rise to the harmonic output voltage, v_{3f}. In this circuit arrangement it is evident that $C(v)$ couples the fundamental and harmonic tuned circuits and affects the tuning of both the fundamental and the harmonic circuits.

The correct tuning of the test circuit, Figure 9, and the values of the fundamental voltage, v_f applied to $C(v)$, and the harmonic output voltage v_{3f}, were determined by means of the loosely coupled circuits L_2C_2 and L_oC_o sharply tuned to the fundamental and harmonic respectively.

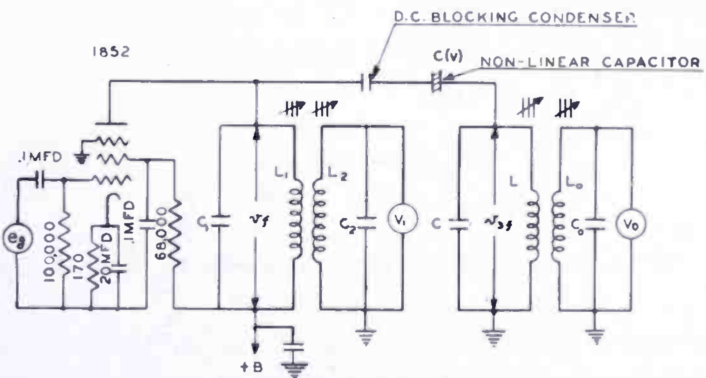


Fig. 9—Circuit used for frequency multiplier tests with non-linear capacitor.

The true efficiency of the non-linear capacitor, $C(v)$, involves both its efficiency as a linear circuit element as well as its efficiency in converting fundamental frequency power into harmonic frequency power. In the following, the figure of merit for $C(v)$ was taken as the ratio of power developed at the third harmonic to fundamental power, and obviously this ratio would be unity for no loss in $C(v)$ and associated circuits. This ratio gives an effective efficiency for $C(v)$ since the circuit losses have not been deducted from the results. This power ratio in terms of convenient measurable quantities was taken to be the following:

$$\frac{\text{Third Harmonic Power}}{\text{Fundamental Power}} = \frac{P_{3f}}{P_f} = \left(\frac{v_{3f}}{v_f}\right)^2 \frac{A_f}{B\,A_{3f}} \tag{1}$$

where v_f, v_{3f} are the voltages as shown in Figure 9.

A_f is the gain at the fundamental frequency as determined by v_f and the fundamental input voltage, e_g, with $C(v)$ in circuit.

A_{3f} is the third harmonic gain as measured at the output circuit, LC, with third harmonic input to the amplifier grid.

The factor B takes into account the fact that the equivalent generator voltage for the third harmonic appears across C_1 in the measurement of A_{3f}, the third harmonic gain as measured at LC.

$$B = \frac{(f_3/f_o)^2}{[1 - (f_3/f_r)]^2} \tag{2}$$

where $f_o = \dfrac{1}{2\pi\sqrt{LC_1}}$: $f_r = \dfrac{1}{2\pi\sqrt{LC}}$: $f_3 =$ third harmonic frequency

The derivation of Equation (1) is readily deduced by recalling that the fundamental power P_f is approximately

$$P_f = \frac{v_f^2}{Z_f} = \frac{g_m v_f^2}{A_f} \tag{3}$$

where Z_f, the fundamental resonant impedance, includes the loading effects of $C(v)$ on the circuit as measured.

An expression similar to Equation (3) in terms of v_{3f} and A_{3f} may be readily deduced by using equivalent series circuit forms[7] and replacing $C(v)$ by a linear capacitor and series resistance so that

$$P_{3f} = g_m v_{3f}^2/B A_{3f} \text{ (approximately)}. \tag{4}$$

The ratio of Equations (4) and (3) yields Equation (1).

As already shown the degree of non-linearity of capacity with field as well as losses greatly diminish at the Curie temperature. This temperature occurs at about room temperature for the 69/31 mixture at which there is a small positive increase in capacity followed by a rather rapid decrease of capacity with field with relatively little increase in power factor as saturation is approached. A sample having an initial capacity of 33 micromicrofarads and 3 per cent power factor when used as a frequency tripler gave the results shown in Figure 10 where the values of the elements of interest are listed in accord with the circuit notation in Figure 9. It was found experimentally

[7] F. E. Terman, RADIO ENGINEERS HANDBOOK (p. 162), McGraw Hill Book Co., New York, N. Y., 1943.

that the maximum third harmonic output as well as the best power ratio was obtained when the coupling of the non-linear capacitor $C(v)$ to the fundamental tuned circuit was large, where about one third the fundamental power appeared as third harmonic power at about optimum as indicated in Figure 10. The results show that it is possible to realize third harmonic voltages almost equal to the fundamental voltage applied, where the efficiency is limited not only by the circuit constants used but also by the increasing losses as the applied voltage is increased.

When the non-linearity of capacity with voltage is increased and the temperature coefficient of capacity at room temperature is reduced by increasing the barium percentage in the barium strontium titanate mixture the losses as a function of voltage also increase considerably

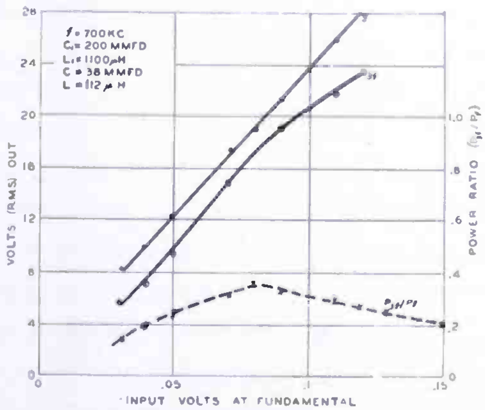


Fig. 10—Titanate 69/31 capacitor as frequency tripler.

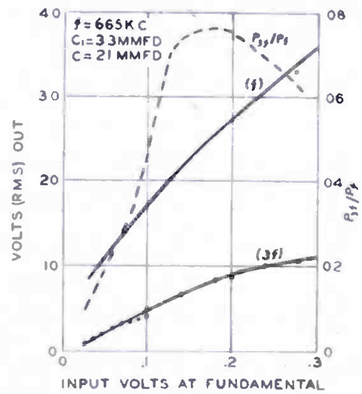


Fig. 11—Titanate 80/20 capacitor as frequency tripler.

as already noted. That the increased losses with increase of applied voltage greatly limit the performance of the dielectric as a frequency multiplier is seen from a comparison of Figures 11 and 10. Figure 11 shows the results for a sample of 80/20 mixture when the effect of the non-linear capacitor on the fundamental circuit is large. For this particular sample the capacity changed 40 to 60 micromicrofarads and the power factor increased from 5 to 15 per cent at about 15 volts applied. Although the third harmonic voltage becomes more nearly equal to the fundamental when the effect of $C(v)$ on the fundamental circuit is great, the efficiency, as judged by the power ratio P_{3f}/P_f is poor when compared with the results obtained for the 69/31 composition, Figure 10.

That the material can also be used as a frequency tripler at 10 megacycles is shown in Figure 12 for another sample of the 80/20 mixture whose capacity increased from 9 to 18 micromicrofarads at

20 volts applied while the power factor increased from 5 to 15 per cent at 12 volts applied.

Although the results as stated apply to thin (1.5-mil) samples, this choice of thickness was purely for convenience in making the various tests.

2. *Frequency Converter*

It is also obvious that such a non-linear dielectric may be used as a frequency converter. Writing out the usual trigonometric expansions for the output current from the non-linear capacitor for an applied voltage consisting of the local oscillator frequency (f_o) and signal frequency (f_s) gives the usual frequency terms of interest namely, $f_o \pm f_s$ when a d-c bias is applied to the non-linear capacitor. The

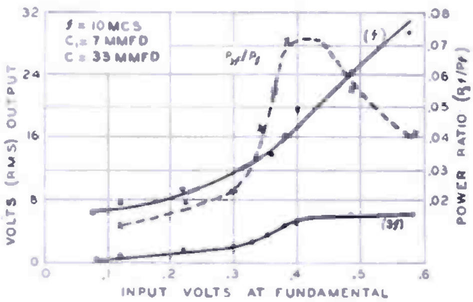


Fig. 12—Performance of 80/20 titanate capacitor as tripler at 10 megacycles.

magnitude of this intermediate frequency will, of course, depend upon the bias value used, as well as on the local oscillator and signal amplitudes.

With no d-c bias present, the change in dielectric constant is independent of direction of radio-frequency flux so that a symmetrical non-linear characteristic results, and as a consequence the intermediate frequency is $2f_o \pm f_s$, the beat between the signal and the second harmonic of the oscillator.

Some measurements were made on these non-linear titanate capacitors as frequency converters in which the frequency translation was accomplished essentially by having the non-linear element common to three tuned circuits tuned to the signal, oscillator and intermediate frequencies. A 30-megacycle signal frequency and 10.7-megacycle intermediate frequency were used, the local oscillator operating at 19.3 megacycles for the straight heterodyning and at 20.35 megacycles for second harmonic operation.

The non-linear titanate capacitors were about 1.5 mils thick with capacity values ranging from 10 to 30 micromicrofarads.

By way of actual performance values, a best noise figure of 15 decibels above thermal noise level (voltage ratio) resulted, from measurements by J. E. Eckert, for a particular 80/20 sample capacitor with second harmonic oscillator operation with an optimum 7 volts oscillator applied. This same sample capacitor gave 21 decibels above thermal noise level for the straight conversion operation with 12 volts oscillator voltage (not optimum) and 6 volts d-c bias applied to the sample. About 20 decibels above thermal noise level resulted for samples of the 69/31 and 90/10 mixtures.

3. *Frequency Modulator*

Another application of the non-linear titanate capacitor is for the purpose of frequency modulating an oscillator. In this application, in order to produce a symmetrical oscillator swing about the center frequency, a d-c bias on the non-linear element is necessary since essentially the non-linear capacitor has a symmetrical non-linear characteristic with voltage.

In an actual qualitative experiment a 5-mil thick titanate capacitor was used to frequency modulate a 40-megacycle oscillator in which the non-linear capacitor was about 30 per cent of the total oscillator tuning capacity. With a 45-volt bias applied to the non-linear capacitor and with 7 volts at 1 kilocycle as modulating frequency a symmetrical oscillator deviation of ± 75 kilocycles was obtained. The modulation characteristics ceased to be approximately linear beyond a 100 kilocycle oscillator frequency deviation.

In order to provide a modulating source of rapid frequency and amplitude changes, the audio output of a broadcast receiver was used as a frequency-modulating signal for the 40 megacycle oscillator. With a frequency-modulation receiver tuned to the 40 megacycle oscillator no audible distortion was noted in the receiver output when the test oscillator deviated no more than ± 75 kilocycles. On the basis of this rough test, the effect of the dielectric "hysteresis" or "viscosity" is negligible for this particular application.

Although no complete non-linear circuit analysis is made in the above described applications, the rough experimental results given may serve to indicate the possibilities in the use of these materials as non-linear circuit elements.

Acknowledgment

Mr. C. Wentworth supplied all the fired titanate bodies and sample capacitor units for the various tests. W. van B. Roberts and D. W. Epstein gave helpful criticisms in their review of this paper.

APPLICATION OF I.C.I. COLOR SYSTEM TO DEVELOPMENT OF ALL-SULFIDE WHITE TELEVISION SCREEN*

By

Austin E. Hardy

Tube Department, RCA Victor Division,
Lancaster, Pa.

Summary—*Increased emphasis on the whiteness of the postwar television tube screen has demanded the use of an objective color-specification system in the development of new phosphors and in the control of the finished tube.*

The I.C.I. color system adopted by the International Commission on Illumination in 1931 has been applied to this problem. In addition to providing accurate control of color, it has made possible the determination of desirable and undesirable color areas and has facilitated the obtaining of the greatest luminosity for any desired white.

Measurements of the relative efficiencies and spectral energy distribution of individual phosphors and phosphor mixtures have been made in a demountable cathode-ray tube with an automatic recording spectroradiometer. Particular emphasis has been placed on the factors that are peculiar to zinc-sulfide and zinc-cadmium-sulfide phosphors.

Introduction

The emphasis placed on the screen whiteness of the postwar television tubes has necessitated increased care in the preparation and evaluation of screen materials and closer control of tube manufacture. In the prewar period, the prevalent factory method for measuring and defining the color of television phosphors was a visual comparison of the tube with variously colored light sources. Visual comparison, in addition to being subjective, was basically inadequate since the colors of even these light sources were undefined.

Very accurate control of color can be obtained by employing some objective color-specification system such as that adopted by the International Commission on Illumination in 1931[1] and commonly referred to as the I.C.I. system[2].

I.C.I. Color System

The development of the I.C.I. system has been adequately presented

* Decimal Classification: R138.313 × R200.

[1] "Report of Commission Internationale de L'Eclairage", Cambridge University Press, 1943.

[2] Sometimes appears in the literature as C.I.E. (Com. Internationale de L'Eclairage).

by various writers[3, 4] and need not be discussed here. Functionally, the system involves the integration over the visible spectrum of the products of visual radiant energy and the three unitary stimuli. The system thus permits the expression of the color sensation in terms of the relative amounts, x, y, z, of these stimuli. Since, in this system, the sum of x, y, and z is always equal to 1, any two co-efficients, by convention x and y, completely specify the color. The integrations could be obtained directly by measuring the radiant energy with filter-phototube combinations which have the I.C.I. primary stimuli response[5]; however, it is not possible to obtain filters and phototubes which match the primary stimuli perfectly at all wavelengths. Hence, this method is limited in its color range. A more fundamental method is to measure the energy-distribution and integrate it by any one of several methods of numerical integrations[6, 7]. The energy-distribution curve can be obtained with a spectroradiometer[8, 9] or from density measurements of spectrograms. This latter method is subject to large error and is mentioned only as a warning against an otherwise obvious approach.

The I.C.I. coordinates, however obtained, may be plotted on a mixture diagram which is simply an x and y plot. The pure-spectrum locus is frequently incorporated on the diagram, but in the development of screens for black-and-white television it serves only to orient the mixture diagram. Of more importance is the Planckian-radiator locus or black-body color line from 3000 degrees Kelvin to ∞. Other helpful adjuncts are the lines of minimum perceptible color-difference (M.P.C.D.) which run more or less parallel to but above and below the Planckian locus, and the isothermal lines which run across the Planckian locus and the M.P.C.D. lines. The isothermal lines do not cross the Planckian locus at 90 degrees because equal linear displacement in any direction on the diagram does not necessarily produce the same magnitude of observed color change. It might be well to mention that there is a uniform chromaticity system of color specification in which

---

[3] A. C. Hardy, HANDBOOK OF COLORIMETRY, Technology Press (Cambridge), 1936.

[4] W. B. Nottingham, "Notes on Photometry, Colorimetry and an Explanation of the Centibel Scale", Radiation Laboratory Report 804, 1945.

[5] B. T. Barnes, "A Four-Filter Photoelectric Colorimeter", *Jour. Opt. Soc. Amer.*, Vol. 29, p. 448, 1939.

[6] Margenau and Murphy, THE MATHEMATICS OF PHYSICS AND CHEMISTRY, D. Van Nostrand Co. Inc., New York, N. Y., 1943.

[7] Bowditch and Null, "Thirty Selected Ordinates", *Jour. Opt. Soc. Amer.*, Vol. 28, p. 500, 1938.

[8] V. K. Zworykin, "An Automatic Recording Spectroradiometer for Cathodoluminescent Materials", *Jour Opt. Soc. Amer.*, Vol. 29, pp. 84-91, 1939.

[9] A. E. Hardy, "A Combination Phosphorometer and Spectroradiometer for Luminescent Materials", *Jour. Electrochem. Soc.*, Vol. 91, pp. 127-146, 1947.

equal linear displacement does represent equal observable color change.[10]

The concept of isothermal lines is somewhat arbitrary and limited in its usefulness. It can apply only to points lying close to the Planckian locus where it is possible to choose a color temperature on the locus which produces a similar color impression[11]. For points not near the Planckian locus there are any number of color temperatures on the locus that are equally different from the color in question, and an equivalent-temperature concept for the point becomes meaningless.

Although the I.C.I. coordinates completely specify a color, more meaning is conveyed by stating that a color is 7000 degres Kelvin + 5 M.P.C.D. rather than by saying that the color is $x = 0.305$, $y = 0.320$.

One extremely useful property of the I.C.I. system is that a straight line between any two points on the mixture diagram represents all the colors that may be obtained by color addition; that is, by mechanically mixing the two luminescent materials. Actually, a straight line is obtained only with materials which have a white body color. If one of the materials shows selective absorption in the visible spectrum, a deviation will develop and appear as a bulge on the otherwise straight line. This deviation assumes considerable importance in the choice of phosphor pairs.

All Sulfide Screens

Certain individual complex phosphors[12, 13] have white light emission, but their efficiency is extremely low. However, efficient physiological whites can be produced by judiciously mixing high-efficiency phosphors of complementary colors.[14]

Because of the relatively higher cathode-ray efficiency of the zinc-cadmium sulfides at current densities of 1 microampere per square centimeter or less, as compared to zinc-beryllium silicates, the former are particularly attractive under present standards for use in television tubes.

An all-sulfide screen is one in which both the yellow and blue components are sulfides. For the sake of completeness one might start the all-sulfide screen development by making a series of silver-activated

---

[10] D. B. Judd, "Maxwell Triangle and a Uniform Chromaticity System", *Jour. Opt. Soc. Amer.*, Vol. 25, p. 24, 1935.

[11] D. B. Judd, "Equivalent Color Temperatures", *Jour. Opt. Soc. Amer.*, Vol. 26, pp. 421-426, 1936.

[12] H. W. Leverenz, "Phosphors Versus the Periodic System", *Proc. I.R.E.*, Vol. 32, No. 5, pp. 256-263, May, 1944.

[13] A. V. Moskvin, "Cathodoluminescence", *Bull. Acad. Sci., U.S.S.R.* (Physics Series 9) pp. 429-460, 1945.

[14] H. W. Leverenz, "Optimum Efficiency Conditions for White Luminescent Screens in Kinescopes", *Jour. Opt. Soc. Amer.*, Vol. 30, No. 7, pp. 309-315, July, 1940.

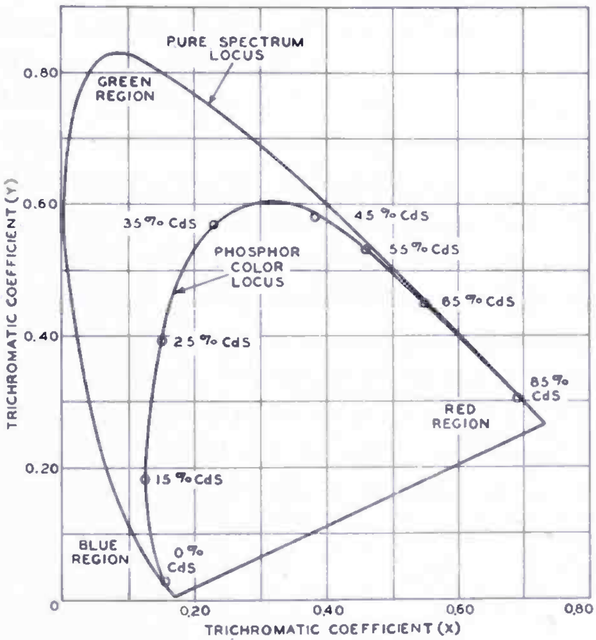


Fig. 1—Mixture diagram showing ZnCdS:Ag phosphor color locus as a function of per cent CdS.

zinc-cadmium-sulfide phosphors containing from 100 per cent zinc sulfide to 70 per cent cadmium sulfide in steps of 10 per cent. The color of each phosphor in the series should then be plotted on the I.C.I. system mixture diagram. A line through these points is the locus for the color as a function of the ratio of zinc sulfide to cadmium sulfide. A typical locus is shown in Figure 1.

In addition to measurements of the color, the visual efficiency[15] and peak emission intensity must also be determined. If all the curves

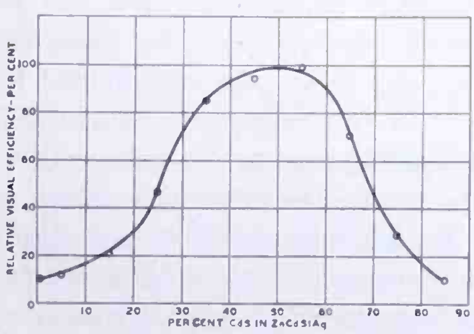


Fig. 2—Variation of relative visual efficiency with per cent CdS in ZnCdS:Ag phosphor.

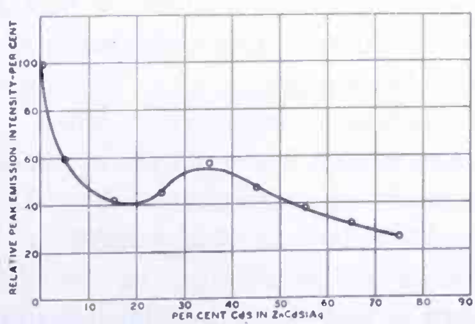


Fig. 3—Variation of relative peak intensity with per cent CdS in ZnCdS:Ag phosphor.

[15] H. W. Leverenz, "Cathodoluminescence as Applied in Television", *RCA REVIEW*, Vol. V, No. 2, pp. 131-175, October, 1940.

of spectral energy distribution were determined under exactly the same condition of excitation and were plotted to scale, then the $\bar{y}$ product summation (Y) of the I.C.I. calculation would represent the relative luminosity or visual efficiency. Plots of relative visual efficiency and peak emission intensity for the various compositions of zinc and cadmium sulfide are shown in Figures 2 and 3.

It will be observed as the cadmium-sulfide content of the phosphor series is increased that, in addition to the shift in fluorescent color, there is also a change in the body color. This shift is brought about by an absorption band at the blue end of the spectrum which broadens regularly toward the red end with increasing cadmium-sulfide content. It is helpful to know the shape of the absorption band at least from 4000Å to as far as it extends in the visible spectrum particularly in the region from 4200Å to 4800Å (Figure 4).

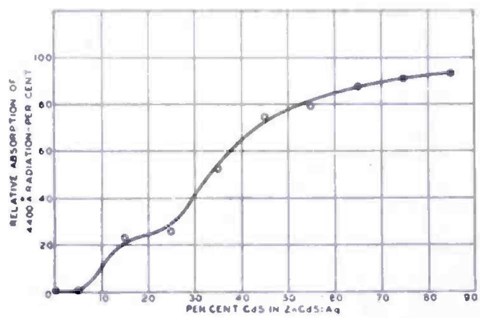


Fig. 4—Variation of relative absorption of 4400 Å with per cent CdS in ZnCdS:Ag phosphor.

Selection of Phosphors

With this information available, it is possible to pick a pair of phosphors which produce a given color temperature with good light output. Assuming that 7000 degrees Kelvin is a desirable white, it can be seen from Figure 5 that a number of straight lines can be drawn through this point each of which will intersect the phosphor color locus at two places. Theoretically, any color pair represented by the points of intersection when mixed in the right proportions will produce 7000 degrees Kelvin. As a practical measure for color control it might seem that color line 2, tangent to the Planckian locus at the 7000 degrees Kelvin point would be desirable, because any slight change in the efficiency of one of the phosphors would shift the color more or less along the Planckian locus. On the other hand, if the color line cuts the locus at an angle, a similar slight change in the efficiency of the phosphors might impart a greenish or purplish tint, depending on the direction of the shift. It has been found that a much greater shift is tolerable along the Planckian locus than across it.

The screen brightness (Y_{80}) of a television tube is related to the intrinsic brightness of the phosphors by the relationship

$$Y_{sc} = \alpha\, Y_Y + \beta\, Y_B \tag{1}$$

where Y_Y and Y_B are the luminosities of the yellow and blue phosphors respectively and α and β are the relative amounts used. Since Y_Y is usually about ten times as great as Y_B, and α and β are equal, most of the visual efficiency of a television screen is due to the yellow component. Figure 2 shows that the maximum visual efficiency in the zinc-cadmium-sulfide series is obtained with a phosphor containing

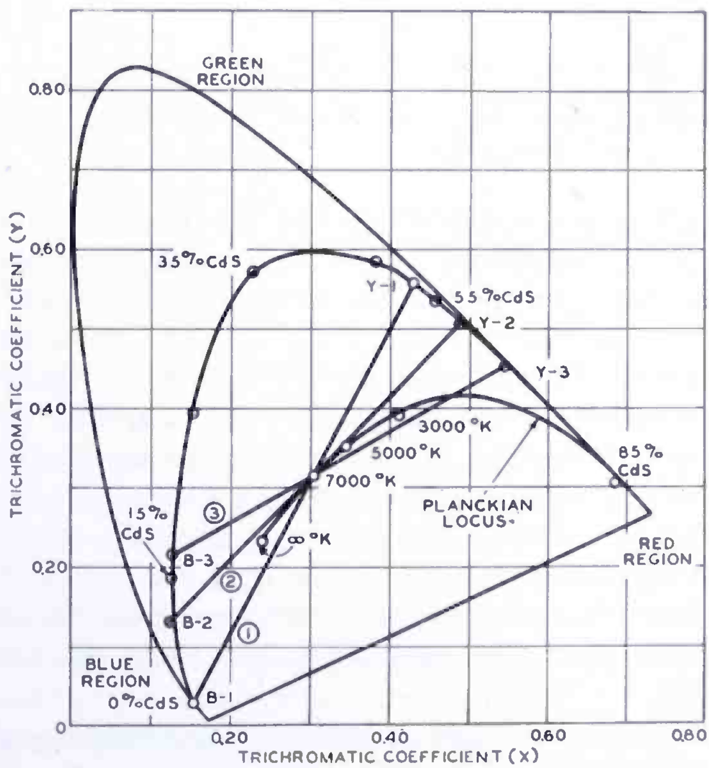


Fig. 5—Mixture diagram showing three phosphor pairs.

45 to 55 per cent cadmium sulfide. It is from this range that the yellow should be chosen. Fortunately, for attaining a color temperature of 7000 degrees Kelvin, it is found that the yellow phosphor required is in this range. Were it not for the selective absorption of the yellow phosphor, the problem would be reduced to making samples *Y*-2 and *B*-2 (Figure 5) and mixing them in the proper amounts, the exact amounts used varying with the peak efficiencies of the individual phosphors.

The following approximation of the relative amounts of blue and yellow needed to make a 7000 degrees Kelvin white may be used if the energy curves are nearly symmetrical and lie almost wholly within the visible spectrum.

$$\alpha = \frac{(100 - \alpha)\, I_B \cdot L_B}{I_Y \cdot L_Y} \tag{2}$$

where

α = per cent yellow to be used,

I_Y = relative peak intensity of yellow phosphor,

L_Y = linear distance from the 7000-degrees-Kelvin point to the yellow end of the color line,

I_B = relative peak intensity of blue phosphor,

L_B = linear distance from the 7000-degrees-Kelvin point to the blue end of the line.

The blue-light absorption of the yellow component effectively displaces the I.C.I. coordinates of the blue phosphor toward the blue green. The displacement increases with increasing cadmium sulfide in the yellow phosphor. Factoring the energy-distribution curve of a given blue phosphor (*B*-2) (Figure 6) by the absorption characteristic of the yellow zinc-cadmium sulfide, *Y*-2, displaces the I.C.I. coordinates of *B*-2 to *B*-2a (Figure 6). If phosphors *B*-2 and *Y*-2 which are supposed to produce 7000 degrees Kelvin were actually used, the colors near the center of the diagram would lie along the center portion of the dotted line *B*-2a—*Y*-2 thus completely missing the Planckian locus. In order to reach 7000 degrees Kelvin using *Y*-2 it is necessary to use a blue phosphor such as *B*-2b which is a deeper blue than *B*-2.

The effective displacement of the blue color due to absorption by the yellow material is obtained by considering the yellow material as a detached filter. In practice the yellow and blue materials are intimately mixed and applied in a thin layer. The magnitude of the absorption is a function of the particle size and the layer thickness. As the magnitude of the blue-light absorption increases, the amount of blue material necessary to produce a particular white has to be increased. Since this increase is effectively a decrease in the amount of yellow material used, the visual efficiency is lowered.

In addition to the effect of the absorption factor, the relative efficiencies of the blue and yellow materials are not constant for changes in excitation conditions, that is, changes in exciting current and voltage. Hence, a phosphor pair which is 7000-degrees-Kelvin at

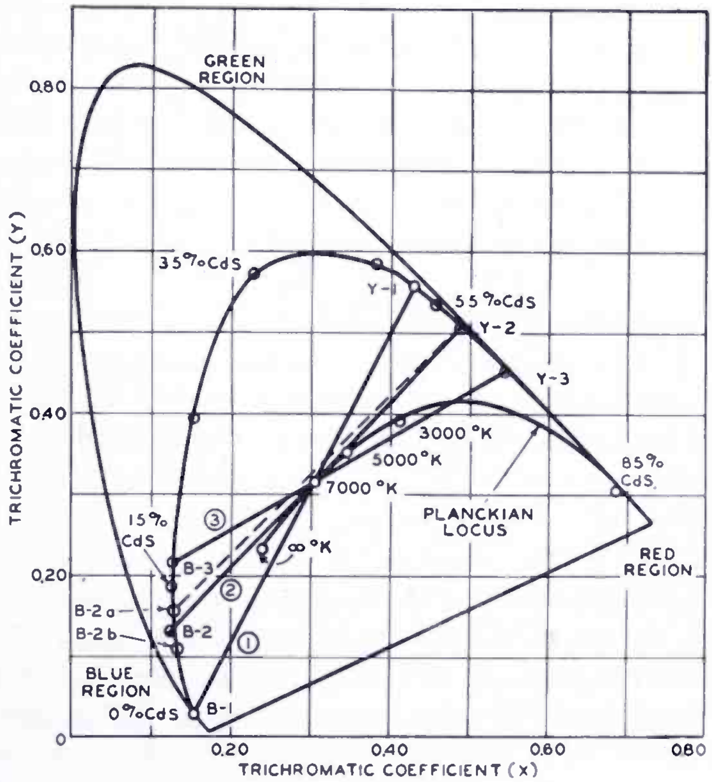


Fig. 6—Mixture diagram showing color displacement of blue phosphor due to absorption.

1 microampere per square centimeter and an accelerating voltage of 8000 volts will shift to a higher color temperature if either the current or voltage is increased, and, conversely, to a lower color temperature if either the current or voltage is decreased. Defocusing a tube will also lower the color temperature since it decreases the current density per luminescent center. The change in color temperature as a function of current and voltage is shown in Figures 7 and 8.

A pair of phosphors for optimum efficiency may be picked from

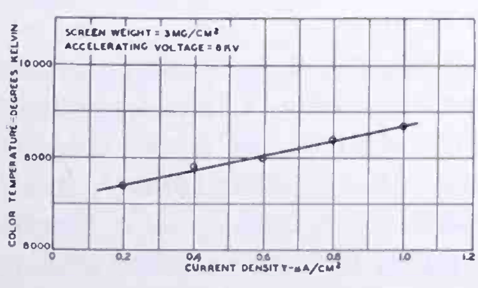


Fig. 7—Effect of change in current density on color temperature.

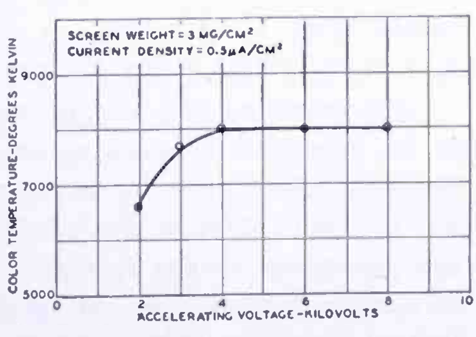


Fig. 8—Effect of change in accelerating voltage on color temperature.

among the three pairs shown in Figure 5 by means of Equations (1) and (2), after weighting them by the blue-light absorption factor. If it is assumed that the relative decrease in absorption due to a thin screen will be nearly constant for the various front-surface absorption values given in Figure 4, then these values may be used.

Equation (2) now becomes

$$\alpha = \frac{(100 - \alpha)\ \gamma\, I_B \cdot L_B}{I_Y \cdot L_Y} \tag{3}$$

where γ is $\dfrac{(100 - \text{the relative absorption})}{100}$ and Equation (1) becomes

$$Y_{sc} = \alpha\, Y_Y + \gamma \beta Y_B \tag{4}$$

The following tabulation shows that phosphor pair No. 1 will produce a brighter 7000-degrees-Kelvin white than either of the others. It should be pointed out, however, that modification of any of the phosphor characteristics might easily reverse the order.

Phosphor Pair	*Per Cent Yellow to Make 7000° K*	*Relative Screen Brightness*
	α	Y_{sc}
1	38	39.3
2	19	21.6
3	9	9.6

Practical Application of Theoretical Values

Developing a 7000-degrees-Kelvin white on paper is simple and direct. Producing the 7000-degrees-Kelvin white in a television tube is more difficult. Factors such as screen weight, particle size, operating conditions, and, of course, the actual color wanted have to be fixed before much can be done. A detailed consideration of these factors is in itself a long study and cannot be presented here.

The principal difficulty in transferring the theoretical composition to the tube lies in resolving the true value of γ. Given a particular phosphor, the value of γ is dependent mainly on the screen thickness but to some extent on the particle size. On the other hand, I_Y and I_B are dependent on the operating conditions. I_Y and I_B as a function of operating conditions can be determined by making tubes with the individual components or by running small patches in a demountable cathode-ray tube. To determine the real value of γ for a particular screen thickness it is necessary to make tubes with a given blend and to measure them under conditions for which I_Y and I_B are known. If two or more compositions are made, an average value for γ can be obtained. This average value is particularly desirable if the initial compositions are quite far from the 7000-degrees-Kelvin color temperature. Once determined, the value of γ can be used in calculating subsequent compositions, barring any change in the body color of the yellow component.

The question whether materials change in color or efficiency during tube processing frequently comes up. The answer is probably that they do not, provided due care is exercised. Even under the best of conditions, however, the answer is not an unqualified negative, since, in a blend, I_B and γ are not easily separated. It can be stated that there is no evidence of a change in color of the components. This does not mean there is no change in the color of the blend, since a change in the efficiency of a component results in a color shift. Comparison between the predicted color and the color obtained in tube production has shown that the color variation from blend to blend can be controlled within $\pm$ 200 degrees Kelvin. This degree of control is sufficient to permit manufacture of carefully compounded blends of a particular color temperature.

Color variation in tube manufacturing seems to be three to four times the blend-to-blend variation. The principal cause of color variation in production, apart from blend variation, is thought to be due to variation in screen weight. Measurements made of screen color as a function of screen weight indicate that a change in screen weight of 10 per cent is equal to a change in color temperature of about 600 degrees Kelvin at an average weight of 4 milligrams per square centimeter.

The author wishes to acknowledge the encouragement of D. H. Wamsley who was among the first to recognize the I.C.I. color system as a valuable aid to screen-color control and development, and to thank J. A. Markoski who synthesized all the phosphors used in this development program.

SPECIAL APPLICATIONS OF ULTRA-HIGH-FREQUENCY WIDE-BAND SWEEP GENERATORS*

By

JOHN A. BAUER

Engineering Products Department, RCA Victor Division,
Camden, N. J.

Summary—Three unusual uses of wide-band frequency-modulated signal generators are described. Instruments suitable for two uses immediately applicable to television receiver development are also shown.

These applications are as follows:—

1. *Radio-frequency impedance measurements;*
2. *Overall frequency response measurements of television receivers;*
3. *Microwave frequency measurements.*

Practicality has already been demonstrated with resulting large laboratory and factory test time savings.

INTRODUCTION

RELATIVELY wide-band frequency-modulated signal generators have been employed for many years in connection with cathode-ray oscilloscopes. The ordinary application has been to use a suitable time base for the examination of frequency response curves in the development and test of all kinds of communications apparatus. Their widest application in the industry prior to World War II, was in production testing and servicing of television receivers. During the war such equipment proved invaluable in speeding the test of radio altimeter receivers, radar receivers, intermediate-frequency strips and many other broad-band radio-frequency circuits.

It is the purpose of this article to present additional uses for this type of generator which are immediately applicable to much of the work being done in development laboratories today. It is hoped that these examples will demonstrate the versatility of such instruments as helpful laboratory tools for quickly determining overall performance of many important components and circuits in television or other broad-band systems. It is also felt that this will stimulate interest in the further refinement of such instruments and their application.

---

* Decimal Classification: R355.913.2 × R200.

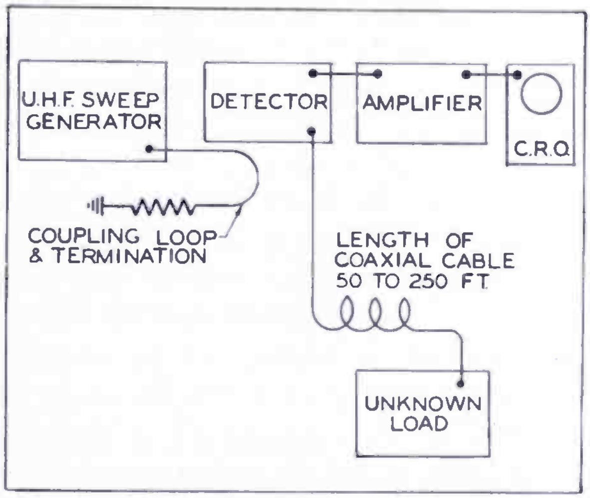


Fig. 1—Block diagram of equipment for radio-frequency impedance measuring method.

Applications

Radio-Frequency Impedance Measurement

This first application has been variously termed as a "Visual Impedance Spectrum Analyzer" or as a "Panoramic Standing Wave Ratio Indicator". A brief mathematical analysis based on Figures 1 and 2 is presented.

By loosely coupling the sweep generator into the sending end of the cable, as shown in Figure 1, generator impedance loading effects are minimized. Under this condition the circuit may be shown schematically as in Figure 2 with the induced voltage, E_{IN}, and the detector impedance, Z_D, across the sending end of a length, L, of

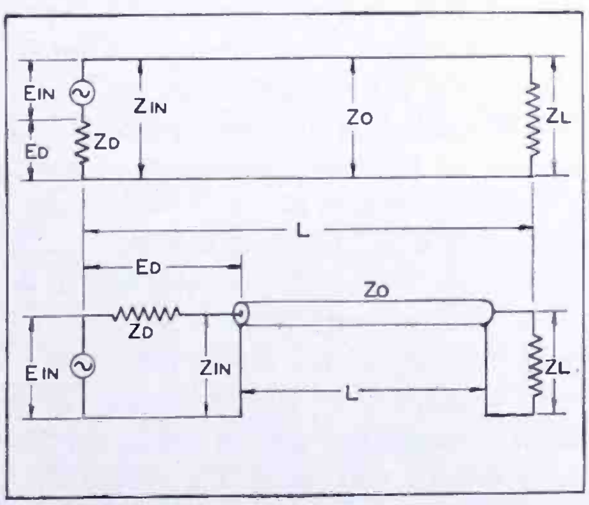


Fig. 2—Schematic of equipment for radio-frequency impedance measuring method.

coaxial line having a specified characteristic impedance, Z_O. A load impedance, Z_L, is connected to the receiving end of the cable.

Examination of Figure 2 will show the following to be true:

$$E_D = E_{IN}\, Z_D/Z_D + Z_{IN} \tag{1}$$

now,

$$Z_{IN} = Z_O \frac{Z_L \cosh \rho L + Z_O \sinh \rho L}{Z_O \cosh \rho L + Z_L \sinh \rho L} \tag{2}$$

then

$$E_D = \frac{E_{IN}\, Z_D}{Z_D + Z_O \dfrac{Z_L \cosh \rho L + Z_O \sinh \rho L}{Z_O \cosh \rho L + Z_L \sinh \rho L}} \tag{3}$$

where: E_{IN} = induced voltage, E_D = voltage across detector,

Z_D = detector impedance, Z_L = load impedance,

Z_O = characteristic impedance of cable, L = length of cable,

Z_{IN} = impedance looking into the cable,

$\rho = (\alpha + j\beta)$, propagation constant of cable.

If Z_L is made equal to Z_O, then Equation (2) reduces to $Z_{IN} = Z_O$ and Equation (3) becomes: $E_D = E_{IN}\, Z_D/Z_D + Z_O$ (4)

This is constant with frequency.

On the other hand, if Z_L is greater or smaller than Z_O, then E_D becomes an oscillating function of frequency. The amplitude of the oscillation depends on the percentage of the incident voltage reflected by Z_L. The period of oscillation depends upon the length of the cable and upon the amount of frequency deviation of the generator. This becomes clearer if one considers the case when total reflection takes place, for example, when a short circuit is placed on the receiving end of the cable. Frequencies for which the cable length is an odd or even number of electrical quarter wave lengths will then be found to occur during the frequency swing of the sweep generator. At a frequency for which the cable appears to be an odd number of electrical quarter wave lengths, highest impedance will be found across the cable input corresponding to a voltage maximum, E_D, across Z_D. Lowest impedance, corresponding to a voltage minimum will similarly be found across the cable input when the cable appears to be an even number of electrical quarter wave lengths long. Continuing with the receiving end of the cable short-circuited the pattern appearing on the cathode-ray oscilloscope will appear as in Figure 3. The number of voltage maxima and minima displayed is directly proportional to the frequency swing and the length of cable. This is true because the number of frequencies for which the line becomes an odd or an even number of electrical quarter wave lengths, increases with both variables.

The short-circuited condition is also convenient for calibration purposes since the peak to peak amplitude of the oscillation can be set equal to ten divisions on the screen of the oscilloscope using the vertical amplifier gain control. This will correspond to 100 per cent reflection. If the cable is then terminated by a typical load such as Z_T, and the peak to peak amplitude over a portion of the frequency

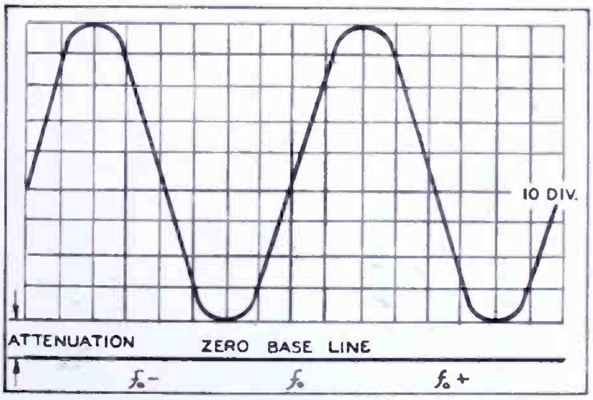


Fig. 3—Cathode-ray tube pattern ($Z_L = 0$; 100 percent reflection).

spectrum, including a voltage maximum and minimum, becomes one division, it is known that the reflected voltage is 10 per cent of the incident voltage, assuming a linear detector, over that range of frequencies. The arrangement can thus be made to display reflection over a wide band of frequencies obtainable in no other more convenient manner.

Additional information is available when the sweep generator is provided with blanking to produce a zero base line as shown in Figure 3. If the cable were actually without losses, the reflected wave would be equal to the incident wave resulting in voltage minima coincident with the zero base line. The distance from the voltage minimum to the base line thus provides a measure of the cable attenuation since, due to attenuation, the reflected voltage is unequal to the incident voltage after traversing down the cable and back. The method could be ideally applied to check long lengths of cable where relative measurements of attenuation compared with a standard of known performance would be satisfactory. Given sufficient voltage level and a well designed detector, measurements of moderate accuracy could easily be made.

It has also been found in practice that internal discontinuities in lengths of cable and the effects of various types of connectors along the length of the cable can be seen when the line is properly terminated.

The immediate applications to which this system has been put include the testing of wide-band terminal and other devices on a transmission line such as television antennas, dummy loads, vestigial side band filters, diplexers (combining aural and visual carriers on one line), triplexers (diplexer combining frequency-modulation carrier on one line), radio-frequency distribution transformers, television

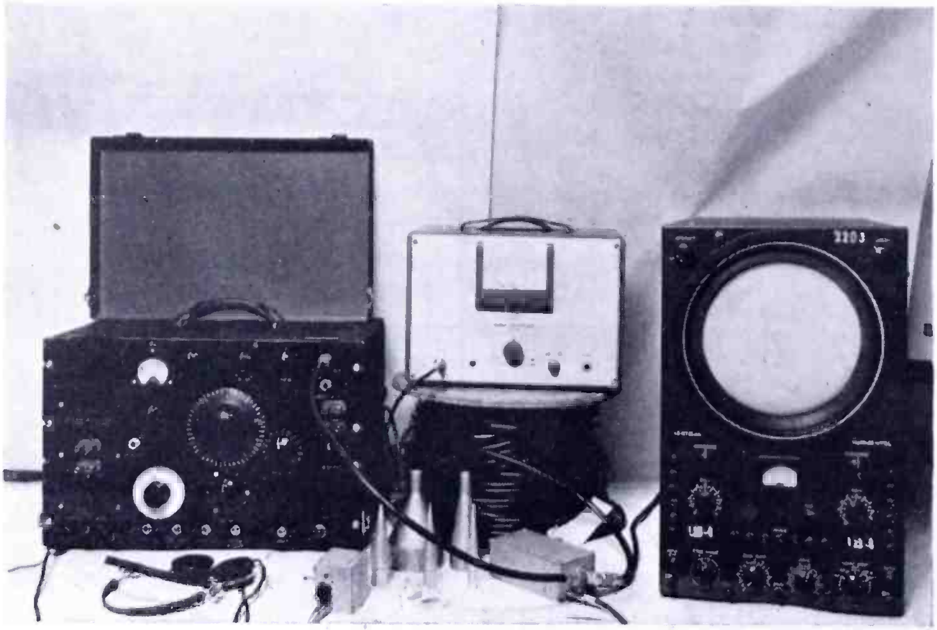


Fig. 4—Panoramic standing wave ratio indicator equipment.

receiver input impedances, etc. Literally months of time, previously expended on standing wave ratio measurements at point-to-point frequencies with a standard slotted measuring line, have been saved by this method. Performance data over a band of frequencies may be seen and the effects of any adjustments on the radio-frequency impedance noted at once.

The panoramic indicator, pictured in Figure 4, is an experimental instrument specifically designed for this application. It is shown in a typical setup with a coil of cable and oscilloscope. The cable is selected to have sufficiently low internal discontinuities as noted above. The small housing with the "T" connector contains the detector bridging the input to the length of line. Accessory taper fittings for measurements feeding into various line sizes are shown.

This unit covers all thirteen television channel frequencies with a sweep 8 to 10 megacycles wide for panoramic standing wave measurements. In addition, a continuous-wave or 60-cycle square wave modulated oscillator with a calibrated attenuator is provided for measuring attenuation through side band filters with the aid of the probe detector and audio voltmeter shown. Two crystal-controlled harmonic generators at 0.25 and 2.5 megacycles allow the oscillator to be set precisely for alignment of sideband notching filters and other precisely tuned circuits in the diplexer and triplexer circuits.

The principal difference between the panoramic indicator and the system described in Figure 1 is the sweep generator output system. While Figure 1 shows the generator loosely coupled to the input of a length of cable this incurs a large loss of signal actually impressed on the line. The panoramic indicator output is connected directly to

Fig. 5—Thirteen channel television sweep generator.

the line input. The resulting increase in signal has not been accompanied by defects traceable to variation of internal impedance of the generator. As far as the analysis of Figure 2 is concerned the internal impedance may as well be considered as part of Z_D.

The set-up does not replace the measuring line altogether except when sufficient familiarity with the method has been attained. It has also been found that correlation with measuring line results quickly establishes this familiarity. Measurements with a slotted line are thereafter supplemented to a large degree with the equipment described.

The thirteen-channel sweep generator, shown in Figure 5, was specifically designed for alignment of television receiver "head end" radio-frequency units. While not as versatile as the panoramic indicator, it has been used as a radio-frequency impedance measuring device with great success. It was the only piece of equipment required to develop a wide-band high-frequency signal distribution system well adapted to video modulated radio-frequency signal distribution from a central transmitter source. Because a detector is built into the unit to monitor the output voltage, a length of cable may be connected directly to the output. An oscilloscope across the detector immediately presents data on impedance connected to the far end of the cable. As a result, distribution transformers having good termination characteristics over the range 40 to 220 megacycles were developed along with signal attenuator panels having satisfactory low reflection. Both these items were used in a video modulated television radio-frequency distribution system for production testing of receivers.

Overall Frequency Response of Television Receivers.

The usual means which have been employed to determine the overall response of broad-band receiving systems such as television receivers require the use of a signal generator capable of being amplitude modulated with a video signal covering a nominal range of 20 cycles to 6 megacycles or as determined by the performance requirements of the receiver. Laboratory signal generators of this kind are not, however, ordinarily available. They must be custom designed and built for the purpose. For example, a special television signal generator was designed for video modulation, out to 8 megacycles, of carrier frequencies in the range 250 to 380 megacycles (airborne television).

To obtain the overall response such a generator was modulated with a video sweep generator which produces a sweep from about 100 kilocycles to 10 megacycles at a 60-cycle rate. The radio-frequency picture carrier frequency for the channel desired was then impressed on the receiver antenna terminal. The overall response could be observed by connecting the vertical amplifier of a wide-band oscilloscope to the video output stage of the equipment under test.

The method to be described obviates the need for a special video modulated generator as mentioned above and makes use of more-readily available equipment such as the thirteen-channel sweep generator. The only additional equipment required is a continuous-wave generator covering the radio-frequency range up to 210.25 megacycles, the picture carrier frequency for channel 13. This is generally available in a moderately well-equipped laboratory.

A more thorough understanding may be gained if one considers a method for checking video response through the second detector, which method eliminates the need for a video sweep generator.

An intermediate-frequency sweep generator signal, 20.75 to 30.75 megacycles, is impressed on the grid of the last video intermediate-frequency amplifier grid. The frequency response of this stage must first be sufficiently broadened by shunting a low resistance across the tuning element to provide a band width of 10 to 12 megacycles centered at the video intermediate-frequency carrier frequency, (25.75 megacycles for many receivers). A continuous-wave generator is also connected to the grid of the last video intermediate-frequency amplifier grid and set for 25.75 megacycles. Levels of these generators may be adjusted for optimum mixing in the second detector. Mixing of these signals in the detector produces a sweep signal of 0 to 5 megacycles which is passed on through the second detector and video amplifier to a cathode follower tube whose input is loaded with capacity to simulate a kinescope input. When an oscilloscope is connected across the cathode resistor a solid pattern will be observed outlining the video response of the receiver. This is subject only to the apparent

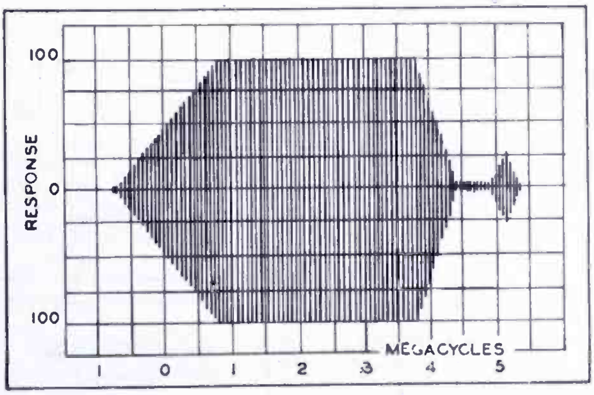


Fig. 6—Beat method used to determine overall response of a television receiver.

change of input impedance in the second detector diode due to loading of the circuit feeding it.

The same type of pattern will now be observed if radio-frequency sweep is fed into the receiver antenna terminals along with a continuous-wave signal. For example a sweep of 40.25 to 50.25 megacycles fed together with a continuous-wave signal of 45.25 megacycles at the antenna terminals will mix with the local oscillator signal of 71 megacycles to produce an intermediate-frequency sweep signal of 20.75 to 30.75 megacycles and a continuous-wave intermediate-frequency signal of 25.75 megacycles. These mix in the second detector (the loading across the last intermediate-frequency stage having been removed) to produce the video sweep. The overall pattern is solid as before and appears as shown in Figure 6. This represents the overall response of the receiver as modified by all the receiver circuits from input to output.

Frequency Measurement

In this application a frequency-modulated oscillator is used in conjunction with a suitable secondary frequency standard, interpolation oscillator, detectors and an oscilloscope to obtain unique advantages in the measurement of frequency. Due to circuit simplicity rapid determination of the approximate frequency of the unknown signal can be made. The system requires a minimum amount of plumbing for microwave work. Cavity wavemeters also are avoided. This reduces to a minimum pulling or loading of the unknown frequency generator with the possibility of throwing it into spurious modes of oscillation.

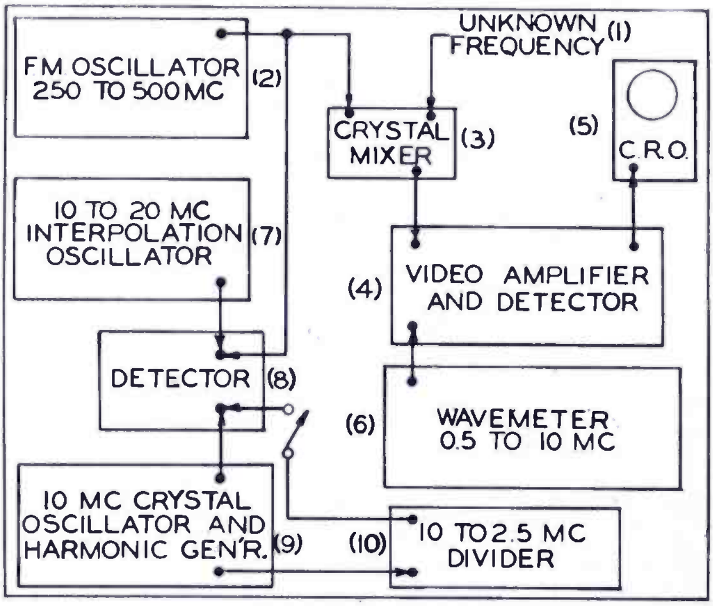


Fig. 7—Block diagram of equipment for frequency measuring system employing frequency modulated oscillator.

The block diagram in Figure 7 outlines the system for measuring unknown continuous-wave frequencies. Assume, for example, that the unknown frequency (1) is 3000 megacycles. If the frequency-modulated oscillator (2) is then adjusted for a swing of plus and minus 1 megacycle at a 60-cycle rate and the center frequency is varied, it will be found that the tenth harmonic of 300 megacycles plus and minus 1 megacycle or 3000 megacycles plus and minus 10 megacycles will beat with the unknown in the crystal mixer (3). Under these conditions the output of the crystal mixer will contain a band of frequencies extending from plus to minus 10 megacycles through zero center. This signal is amplified by the video amplifier and detector (4).

The detected voltage envelope is displayed on the oscilloscope (5). The wavemeter (6) is used to determine the order of harmonic above the fundamental generated by the frequency-modulated oscillator which beats with the unknown. Measuring the extent of the resultant frequency deviation appearing in the output of the crystal mixer will yield this information. This may be done by an absorption dip or mixed-in marker signal from the wavemeter moved out to the ends of the envelope appearing on the oscilloscope. If the markers coincide with the ends of the envelope when the wavemeter is adjusted to 9 megacycles it is known that the unknown frequency is beating with the ninth harmonic of the frequency-modulated oscillator when set at 300 megacycles with plus and minus 1 megacycle swing, or 2700 megacycles plus and minus 9 megacycles.

Unknown frequencies as high as 10,000 megacycles have been conveniently identified by these means when the frequency-modulated oscillator was adjusted to a center frequency of 500 megacycles with a swing of plus and minus 0.5 megacycle. The twentieth harmonic of the frequency-modulated oscillator was then beating with the unknown and could be identified with wavemeter (6) set to 10 megacycles.

After this approximate determination the deviation of the frequency-modulated generator is reduced to zero and the center frequency readjusted to zero beat with the unknown. If the dial of the generator has 500 divisions and the frequency distribution is linear, the frequency of the unknown signal will be directly indicated to within 0.5 megacycle times the order of the harmonic or 1 part in 500.

It now remains to establish the frequency of the oscillator more accurately while zero beating with the unknown. This is done by conventional means. The 10-megacycle crystal oscillator and harmonic generator (9) is fed into the detector (8) along with some signal from the oscillator. A train of standard frequencies is then available at 10-megacycle intervals to establish this frequency. When the frequency of the oscillator is displaced from the 10 megacycle intervals the interpolation oscillator (7) is used. For example, if the frequency happens to be at 333 megacycles the output of the detector will contain . . . . 23, 13, 3, 7, 17, 27 . . . . megacycles. From the calibration of the oscillator the frequency is already known within 0.5 megacycle and one need only measure the displacement 3 megacycles above the nearest standardized 10-megacycle interval, 330 megacycles. This can be done as well by measuring the displacement, 13 megacycles, above the nearest standardized 10-megacycle interval but one, 320 megacycles. This procedure simplifies the design requirements of the interpolation oscillator by restricting the frequency range from 10 to 20 megacycles rather than 0 to 10 megacycles.

If the dial of the interpolation oscillator has 5000 divisions and the 10 to 20 megacycle range is linear with frequency one can easily read to within ½ division or 1 kilocycle above the lowest frequency interval of 250 megacycles from the harmonic generator. This represents a first error of ± 0.0004 per cent. A second error of ± 0.0002 per cent must be allowed for the crystal. A third error of ± 0.0004 per cent or 1000 cycles (200 cycles is practical) may be allowed for adjusting zero beats between the unknown signal and the frequency-modulated oscillator (deviation set to zero) as well as between the signal and the interpolation oscillator. An accuracy of ± 0.001 per cent can thus be realized. It is expected that while the first error is cut in half at the high frequency end of the oscillator the third error will increase correspondingly due to added difficulty in maintaining zero beat. However, the total accuracy of ± 0.001 per cent should prove practicable over the entire range.

The interpolation oscillator must be temperature controlled. The high frequency divider (10) may also be available to calibrate the interpolation-oscillator scales at 2.5 megacycle intervals or closer. The 10-megacycle harmonic generator and the divider are useful also to maintain calibration of the oscillator.

The only plumbing required for a frequency range of 250 to 10,000 megacycles is a small number of crystal mixer mounts, item (3) in Figure 7. There does not seem to be any obvious limitation on this system for applications above 10,000 megacycles.

Conclusions

The engineer concerned with the development of wide-band ultra-high-frequency circuits will be amply repaid by taking the time to become familiar with the principles outlined in the first application. For example, the alignment of a vestigial side band filter which was formerly a matter of weeks is now reduced to days.

The second application will be of immediate interest because of the lack of suitable television modulated signal generators for laboratory development work.

From present knowledge of the field, it is not believed that there is any equipment to compare with that outlined in the third application. No equipment appears to give such ready results with comparable accuracy over such a large portion of the radio-frequency spectrum with the simplest kind of plumbing.

It is probable that the first and last applications could be further developed and refinements beyond those outlined here brought to light.

Members or former members of the Test and Measurements Section in Camden, N. J. were responsible in each case for some phase of the work described. W. J. Frantz suggested the first application, the visual impedance spectrum analyzer, and presented the brief mathematical analysis on November 6, 1944. A. J. Grange developed the panoramic standing wave ratio indicator equipment as applied to television work. The second application was described by Edw. S. Clammer early in May 1947. The third application was described in a patent disclosure May 17, 1944, by M. J. Ackerman, Edw. S. Clammer and G. Barton.

RCA TECHNICAL PAPERS†

Second Quarter, 1947

Any requests for copies of papers listed herein should be addressed to the publication to which credited.

"An Analysis of Modern Antennas for FM and Television Reception", M. Kaufman, *RCA Rad. Serv. News* (June-July) 1947

"Cathode Ray Unit Speeds Industrial Applications", P. A. Greenmeyer, *Electronic Ind.* (April) 1947

"Circularly-Polarized Omnidirectional Antenna", G. H. Brown and O. M. Woodward, Jr., *RCA REVIEW* (June) 1947

"The Clamp Circuit — Part II", C. L. Townsend, *Broad. Eng. Jour.* (April) 1947

"Coaxial Tantalum Cylinder Cathode for Continuous-Wave Magnetrons", R. L. Jepsen, *RCA REVIEW* (June) 1947

"Comparator for Coaxial Line Adjustments", O. M. Woodward, Jr., *Electronics* (April) 1947

"Continuous Soldering of Small Motor Rotors Using High-Frequency Heat", W. L. Tesch and P. A. Greenmeyer, *Materials and Methods* (June) 1947

"Criteria for Diversity Receiver Design", W. Lyons, *RCA REVIEW* (June) 1947

"Demonstration, with the Electron Microscope, of a Nucleus in Bacillus Mycoides Grown in a Nitrogen-Free Medium", R. F. Baker (Coauthor), *Jour. Bacteriology* (May) 1947

"Edge Gluing Economy and Equipment—A Tentative Analysis", E. S. Winlund, *Wood Products* (June and July) 1947

"The Electron Mechanics of Induction Acceleration", J. A. Rajchman and W. H. Cherry, *Jour. Frank. Inst.*, Part I (April) 1947
Part II (May) 1947

"Electronic Attenuators", F. W. Smith, Jr. (Coauthor), *Communications* (May) 1947

"Explanation of the Ratio Detector as an Aid in FM Servicing —Part II", J. A. Cornell, *RCA Rad. Serv. News* (June-July) 1947

"Film Projectors for Television", R. V. Little, Jr., *Inter. Project.* (May) 1947

"A High-Quality Recording Power Amplifier", K. Singer, *Jour. Soc. Mot. Pic. Eng.* (June) 1947

† Report all corrections or additions to RCA REVIEW, Radio Corporation of America, RCA Laboratories Division, Princeton, N. J.

"Input Admittance of Receiving Tubes", *RCA Application Note AN-118,* RCA Tube Department, Harrison, N. J. (April 15) . 1947

"Input Impedance of a Folded Dipole", W. van B. Roberts, *RCA REVIEW* (June) 1947

"The Mechanism of the Luminescence of Solids", F. E. Williams (Coauthor), *Jour. Chem. Phys.* (May) 1947

"Miniature Tubes in War and Peace", N. H. Green, *RCA REVIEW* (June) 1947

"Multiplier Photo-Tube Characteristics: Application to Low Light Levels", R. W. Engstrom, *Jour. Opt. Soc. Amer.* (June) 1947

"New Techniques in Synchronizing Signal Generators", E. Schoenfeld, W. Brown and W. Milwitt, *RCA REVIEW* (June) 1947

"A New 50 KW AM Transmitter", W. L. Lyndon, *Broadcast News* (June) 1947

"A Non-Directional Antenna for Mobile Field Strength Measurement in the FM Band", B. W. Robins, *Broadcast News* (June) 1947

"Operation of the RCA-6SB7-Y Converter", *RCA Application Note AN-120,* RCA Tube Department, Harrison, N. J. (June 16) 1947

"The Outlook for the Radio Industry", David Sarnoff, pamphlet, RCA Department of Information, New York, N. Y. (June 12) 1947

"Phase-Front Plotter for Centimeter Waves", H. Iams, *RCA REVIEW* (June) 1947

"Precision Device for Measurement of Pulse Width and Pulse Slope", H. L. Morrison, *RCA REVIEW* (June) 1947

"Radar for Merchant Marine Service", F. E. Spaulding, Jr., *RCA REVIEW* (June) 1947

"Radiation Angle Variations from Ionosphere Measurements", H. E. Hallborg and S. Goldman, *RCA REVIEW* (June) 1947

"Radio — Frequency vs. Utility", E. A. Laport, *Electronics Buyers' Guide* (June, Mid-month) 1947

"The Radio Mike", J. L. Hathaway and R. Kennedy, *RCA REVIEW* (June) 1947

"The Ratio Detector", S. W. Seeley and J. Avins, *RCA REVIEW* (June) 1947

"A Review of Criteria for Broadcast Studio Design", H. M. Gurin and G. M. Nixon, *Jour. Acous. Soc. Amer.* (May) 1947

"The Selectron — A Tube for Selective Electrostatic Storage", J. Rajchman, *RCA Labs. Div. News* (June) 1947

"Simplified Television for Industry", R. E. Barrett and M. M. Goodman, *Electronics* (June) 1947

"Single Crystal Electron Diffraction by Micro-Crystalline Materials", N. Davidson and J. Hillier, *Jour. Appl. Phys.* (June) . 1947

"Some New Aspects of Germanate and Fluoride Phosphors", F. E. Williams, *Jour. Opt. Soc. Amer.* (April) 1947

"Stabilized Magnetron for Beacon Service", J. S. Donal, Jr., C. L. Cuccia, B. B. Brown, C. P. Vogel and W. J. Dodds, *RCA REVIEW* (June) 1947

"A Study with the High-Voltage Electron Microscope, of the Endospore and Life Cycle of Bacillus Mycoides", R. F. Baker and J. Hillier (Coauthors), *Journal of Bacteriology* (May) .. 1947

"Teleran", D. H. Ewing and R. W. K. Smith, *Airways Radio Journal* (April and June) 1947
Reprinted from *RCA REVIEW* (December) 1946

"Television Antenna Installations Giving Multiple Receiver Outlets", R. J. Ehret, *Tele-Tech* (June) 1947

"Television Today and Its Problems — 1946", A. N. Goldsmith, *Inter. Project.* (May) 1947
Reprinted from TELEVISION, Vol. IV (January) 1947

"The Theory and Design of Speech Clipping Circuits", M. H. Dean, *Tele-Tech* (May) 1947

"Use of the 2E24 and 2E26 at 162 Megacycles", *RCA Application Note AN-119*, RCA Tube Department, Harrison, N. J. (May 15) 1947

"Use of the 6BA6 and 6BE6 Miniature Tubes in FM Receivers", *RCA Application Note AN-121*, RCA Tube Department, Harrison, N. J. (June 16) 1947

NOTE—Omissions or errors in these listings will be corrected in the yearly index.

Correction:

On page 67 of the March 1947 issue, a typographical error appeared in Equation (35). The equation as printed was:

$$Q_1 \geqq 10/0.33 = 333$$

This should have read: $Q_1 \geqq 10/.03 = 333.$

AUTHORS

JOHN A. BAUER attended Drexel Institute Evening School in Philadelphia, Pa., while employed in the Long Lines Department of the A. T. & T. Co. He received his E.E. degree in 1933 while associated with the Philco Corporation. As factory engineer he initiated television receiver manufacturing tests on a pilot production basis. Since 1940 he has been with RCA Victor Division, Camden, N. J., as design engineer, group leader and a supervisor in the Test and Measurements Section. Mr. Bauer is a Member of Institute of Radio Engineers.

WILLIAM H. CHERRY received the B.S. degree in physics from the Massachusetts Institute of Technology in 1941. He joined the research staff of RCA Manufacturing Co. at Harrison, N. J., in August, 1941, and transferred to RCA Laboratories Division at Princeton, N. J., in the following year. He is at present working in the television group, and is also studying graduate physics at Princeton University. Mr. Cherry is a member of Sigma Xi.

HUGH L. DONLEY received the B.S. degree from Hobart College in 1930, the M.Sc. and the Ph.D. degrees from Brown University in 1932 and 1935 respectively. From 1930 until 1933 he was assistant in Physics at Brown University, and from 1933 to 1934 he was a Fellow in Physics at Brown University. From 1935 to 1942 he was in the research division of RCA Manufacturing Company at Camden, N. J. Since 1942 he has been with RCA Laboratories Division, Princeton, N. J. Dr. Donley is a member of Phi Beta Kappa, Sigma Xi, and a Senior Member of the Institute of Radio Engineers.

LAWRENCE J. GIACOLETTO received the B.S. degree in Electrical Engineering from Rose Polytechnic Institute, Terre Haute, Ind. in 1938 and the M.S. degree in Physics from Iowa State University in 1939, while holding an appointment as research assistant there. From 1939 to 1941 he was a Teaching Fellow in the Department of Electrical Engineering at the University of Michigan. He was associated with the Collins Radio Company during the summers of 1937 and 1938 and with the Bell Telephone Laboratories in 1940. From 1941 to 1946 he served with the Signal Corps in various laboratory and field (in Japan) assignments concerned with development activities in the field of radio, navigational and meteorological direction finding equipment. He returned to inactive status as a Major in the Signal Corps Reserves in May, 1946 and since then has been serving as a research engineer with RCA Laboratories Division, Princeton, N. J. Mr. Giacoletto is a member of the American Association for the Advancement of Science, Tau Beta Phi, Gamma Alpha, Iota Alpha, Phi Kappa Phi, and Sigma Xi.

AUSTIN E. HARDY graduated from the University of New Hampshire in 1943 with the degree B.S. in chemistry, and joined the staff of the RCA Victor Division at Lancaster, Pa. In 1946 he was awarded the Young Authors prize by the Electrochemical Society for his paper on "The Photoconductivity of Zinc-Cadmium Sulfide as Measured with the Cathode-Ray Oscillograph." This paper was the result of an original approach to a performance evaluation problem on phosphors for radar cathode-ray tubes. Mr. Hardy is in charge of developing methods for phosphor evaluation.

G. ROSS KILGORE received the B.S. degree in electrical engineering in 1928 from the University of Nebraska and the M.Sc. in electrical engineering in 1931 from the University of Pittsburgh. He entered the employ of the Westinghouse Electric Corporation in East Pittsburgh in 1928. In 1934 he was employed by the RCA Radiotron Company in Harrison, N. J., to work on research and development of ultra-high-frequency and super-high-frequency transmitting and receiving tubes. From 1942 through 1946, he was engaged as a research scientist at RCA Laboratories Division, Princeton, N. J. Recently he joined the Evans Signal Laboratory in Belmar, N. J. Mr. Kilgore is a member of Sigma Tau, Sigma Xi, Pi Mu Epsilon, and a Senior Member of the Institute of Radio Engineers.

ROBERT G. KREER was graduated from Rutgers University in 1935 with a degree of B.S. in Electrical Engineering. He was employed by the Weston Electrical Instrument Corporation and in April, 1936, was assigned to the office of that company's Philadelphia representative where he was engaged as Sales Engineer. Mr. Kreer entered active duty with the Signal Corps in March of 1941 and served for the most part in the Army Communications Service until his separation in 1946 in the rank of Lieutenant Colonel. He joined the staff of RCA Communications Inc. where he has the position of Traffic Engineer.

OTTO H. SCHADE graduated from the Reform-Real-Gymnasium, Halle, Germany, in 1922. From 1922 to 1924 he was with the Telephonfabrik A. G. vorm. J. Berliner, Berlin and Dusseldorf; from 1924 to 1925, in charge of the laboratory in the radio manufacturing company "Ratag" in Berlin; and from 1926 to 1931, in the engineering department of the Atwater Kent Manufacturing Company. Since 1931 he has been with the Tube Department, RCA Victor Division at Harrison, N. J. He received a Modern Pioneer Award from the Radio Manufacturers Association in 1940. Mr. Schade is a Senior Member of the Institute of Radio Engineers.

SIDNEY SPARKS received his B.S. degree in Electrical Engineering from The Catholic University of America in Washington, D. C., in 1927. Prior to the war, Mr. Sparks was Division Sales Manager for the Western Union Telegraph Company in New York. During the war, he was assigned to the Office of the Chief Signal Officer in Washington, D. C., and in September 1945, joined the staff of RCA Communications, Inc., as Traffic Manager. He was elected a Vice President in April 1946