IRE TRANSACTIONS

ON BROADCASTING

Volume BC-6 SEPTEMBER, 1960 Number 3

TABLE OF CONTENTS

PUBLISHED BY THE

PROFESSIONAL GROUP ON BROADCASTING

IRE PROFESSIONAL GROUP ON BROADCASTING

The Professional Group on Broadcasting is an organization, within the framework of the IRE, of members with principal professional interest in Broadcasting. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of prescribed fee.

Annual Fee: $2.00

IRE TRANSACTIONS®

on Broadcasting

Published by the Institute of Radio Engineers, Inc., for the Professional Group on Broadcasting at 1 East 79th Street, New York 21, New York. Responsibility for the contents rests upon the authors, and not upon the IRE, the Group or its members. Individual copies available for sale to IRE-PGBC members at $0.65, to IRE members at $0.95, and to non-members at $1.95

ON THIS ISSUE.......

This issue of the PGB transactions is designed to appeal to practically everyone. It includes comments on FM broadcast standards by O. W. B. Reed and the engineering aspects of an educational FM network by Roger Peterson, both in keeping with the current rising interest in FM broadcasting. Next comes a discussion of the use of phase equalizers to improve television transmitting system transient response by J. K. McDonald of the Canadian Broadcasting Corporation. Then John Wentworth tells us about tunnel diodes and some of their potential uses. Finally, we have two European papers. One, by J. Davidse, is concerned with the transmission of color television signals and gives some ideas of the adoption of NTSC principles to the 625 line system, and gives as well a little different approach to tritanopia characteristics of the eye as applied to color television. The last paper, by H. F. Mayer and F. Bath, is on stereophonic braodcasting using pulse amplitude modulation. These last two papers are republished by permission of the authors.

The next issue of PGB transactions will be primarily taken up with the fall symposium papers and will be out around the first of the year. This gives just enough time for all of the potential authors to submit papers for the March 1961 issue. We have had many varied comments from broadcasters on what should be in the transactions and we would like more, so don't hesitate to write letters to the editor suggesting ideas and papers.

One thing that has been suggested is to have a "Hints and Kinks" column from broadcasters to other broadcasters--how about it? If you like this idea, or have others, write to us or better yet, tell us at the fall symposium on September 23 and 24 at the Willard Hotel in Washington, D.C. See you there.

Editor

COMMENTS CONCERNING FM BROADCAST STANDARDS
OF THE FEDERAL COMMUNICATIONS COMMISSION

O. W. B. Reed, Jr.
Jansky & Bailey, Inc.
Washington, D. C.

There are indications at present that the Federal Communications Commission is seriously considering revisions of its Rules concerning FM broadcast stations. The Technical Standards contained as a part of these Rules specify engineering standards of allocation which relate to FM station coverage, determination of interference between stations, the location of transmitters and other technical factors. Considerable incentive to revision of the FM Technical Standards has been occasioned by the availability of new field strength measurements made in connection with the program of the Television Allocations Study Organization. The results of this measurements program will probably yield new field strength versus distance curves for FM broadcasting.

The Association of Federal Communications Commission Consulting Engineers has recently appointed a committee to consider the matter of revising FM Standards and co-operate with the working groups of the Federal Communications Commission engaged in this effort. The FM Rules issued in September 1945 have been revised in minor respects only over the intervening years and the current undertaking will represent the first complete revision. It is interesting to note that in June 1947, the Federal Communications Commission engineering staff developed a Table of Assignments for FM stations which was used as an adjunct to the Rules until its abolition in August 1958. Since this latter date procedures similar to those prevailing in standard broadcasting for the selection of available frequencies have been followed.

Today there are 967 authorizations outstanding in the United States for FM broadcast facilities. These include either existing licenses or outstanding construction permits. At present there are 704 Class B FM, 97 Class A FM, and 166 noncommercial educational FM authorities. The number of applications presently pending in these various categories totals an additional 86. With the increased growth in FM interest which these figures indicate the Federal Communications Commission is apparently concerned that new engineering factors which have come to light in the past 10 or 12 years be taken into account as this type of broadcasting continues to develop. The pending Notice of Inquiry concerning the multiplex services now provided for under the Subsidiary Communications Authorization program and the Notice of Inquiry into Stereo developments are not being directly associated with the immediate task of revising the present Standards. Industry comments in the Stereo matter are presently due in early December 1959. It is possible that these matters will take somewhat longer to include in the Rules and Standards.

The present FM Standards when subjected to individual review on a part-by-part basis indicate that items such as the following may be considered:

1. Pertinent ground-wave contour values for urban and rural coverage
2. Field strength necessary over the principal city to be served
3. Prediction of service areas
4. Determination of population served
5. Interference standards
6. Field strength measurements
7. Transmitter siting
8. Antenna system
9. Transmitter and associated technical equipment
10. FM propagation curves

These 10 features of the Technical Standards will certainly be among those considered for review.

Since Standards of good engineering practice are developed to provide the guide lines for the development of FM within a framework of Rules concerning the number of channels for the various categories of stations they are bound to be a function in some respects of the Rules themselves. Corollary consideration of Allocation Rules is therefore necessary as a part of the deliberations which will occur. For example, one might ask, with the benefit of several years in development behind us the following questions:

1. Is there presently a need to separate categories of stations, namely, Class A FM stations which operate with limited power and height and Class B stations which employ somewhat greater power and height?

2. Is it necessary that the United States be divided into Area I and Area II where different power and height considerations govern?

3. Are the existing power and height limitations for the several classes of stations good features to be retained or should they be modified in some respects?

4. How should the provisions with respect to Subsidiary Communications Authorization be modernized in light of technical developments?

These questions involve policy determinations which conversely may be viewed as influenced by Standards as the state of the art progresses.

There are policy questions pending which relate to matters outside the field of the provisions of the present FM rules and Standards. For example, there are no requirements in either the FM Rules or the AM Rules at present which relate to Stereo broadcasting. However, AM and FM Stereo operations are presently carried on under the authorities of licenses presently in effect to demonstrate the capability of Stereo broadcasting. Similar demonstrations are being made by pairs of licensed FM stations. These are matters with which the industry is tremendously concerned at present and upon which it is logical to expect further explorations of this nature. Such explorations, for example, are going on in program relaying and other situations for which specific authorities have been obtained.

To return to several of the specific Standards previously mentioned, the introduction of revised FM coverage curves and curves for predicting tropospheric signals will certainly be of interest. The question of appropriate balance between station coverage and provision for a maximum number of stations within the frequency assignments available must of necessity be weighed. Desires for high effective radiated power and greater antenna heights are of necessity modified at present by the shortage of channels in some east coast and west coast areas.

The Federal Communications Commission has found it necessary to hold hearings in recent years between applicants where the present Standards indicate the availability of just a single remaining assignment.

Provisions with respect to FM antenna systems also will be of importance. In recent years a proposal was filed with the Federal Communications Commission that FM broadcast stations be allowed to operate with either horizontal or vertical polarization. This petition was subsequently denied. However, the present Standards indicate that while horizontal polarization shall be the standard for FM, circular or eliptical polarization may be employed if desired. Thus, within these present Standards a signal having some vertical component may be provided and certain experiments have been carried out along this line. Results indicate that this can be beneficial in consideration of the problem of providing adequate field strength to automobile receivers which conventionally operate using a vertical antenna. The present provision of the Standards in this respect requires merely that the supplemental vertically polarized effective radiated power shall in no event exceed the effective radiated power authorized to the licensee.

The provisions relating to Interference Standards, and the required ratio of desired-to-undesired signals for the same channel and channels up to 800 kilocycles removed in frequency from this channel, are certainly pertinent to the consideration. The characteristics of FM receivers as regards their selectivity will undoubtedly be factored into such deliberations. In this connection it may be recalled that in 1945 it was considered that stations could operate satisfactorily in the same metropolitan area when separated by only 400 kilocycles. As stations increased in number however it very quickly became necessary to revise this Standard to provide for 800-kilocycle separation.

The need for revising the FM Rules and Standards is apparent to all who have worked in the field of FM in recent years. A general review of the possible scope of the revision has been indicated herein. The coming pattern for FM broadcasting for the next several years will be influenced to some extent by the deliberations now under way at the Federal Communications Commission. The aforementioned topics of interest are merely typical of those which will be given scrutiny.

A thorough review of all matters of possible concern would seem to be warranted by those broadcasters who are in the field today. Their experience should be used to aid the efforts and achievements of those who will continue to add more to the stature of FM broadcasting tomorrow.

THE ENGINEERING ASPECTS OF AN EDUCATIONAL FM NETWORK

Roger E. Peterson
Jansky & Bailey, Inc.
Washington, D. C.

Summary

FM radio relay techniques have been demonstrated since the early days of FM broadcasting. Several groups of commercial FM stations have formed small, loosely knit networks with varying degrees of technical performance and acceptance by the public. Educational institutions have been slow to realize tangible benefits from FM broadcasting.

The current resurgence of FM broadcasting in general and a new evaluation of the potential of aural broadcasting for education has prompted detailed plans for an extensive FM educational network. Educators recognize their need to know more; to exchange information and ideas with each other and with the faculties of other institutions; in brief to communicate. It's a special demand, sophisticated and vital to the educational system of today. It has not been met with commercial radio and television. The educational FM radio network described herein, is designed to meet the needs of the educational system of today.

Multiplexing techniques would be used to enhance program handling capability and also provide remote control features allowing important economies in operation.

Introduction

A plan has been conceived by the WGBH Educational Foundation for the establishment of an FM radio relay network throughout the Northeastern United States to provide a unique aural service to educational institutions within this area. The network could serve all major cities from Montreal, Canada to Washington, D. C. and eastward to Boston, Massachusetts. At least 11 stations would be required to provide the desired service, some of which are now existing, being owned and operated by local educational institutions. Interconnection with other educational networks in Wisconsin, Ohio, Michigan, Pennsylvania, Virginia, South Carolina, and New York would be possible as an extension of the system. The engineering aspects of the proposed network are discussed in this paper.

The details of the proposed network presented herein are the results of extensive investigations by WGBH and Jansky & Bailey, Inc., and are based on specifications presently available for the most part in production equipment supplied by various manufacturers. Since some of the engineering details cannot be defined wholly before implementation of the first portions of the plan, certain of the system specifications must be regarded as tentative.

System Philosophy

The basic concept of the proposed plan has been to provide an aural service of maximum usefulness to the participating organizations within the service area, at a minimum operating cost. Considering the concept of maximum usefulness, the first requirement is to provide an adequate level of signal throughout the area to be served. The size of the area will, of course, require multiple stations, located strategically throughout the service area to provide adequate signal strength and to minimize holes or signal shadows due to hilly terrain. The frequency allocations structure of the area must be carefully analyzed to select frequencies for the stations which will encounter minimum interference from other authorized stations.

For an educational program to be of maximum usefulness it must be aired at a time convenient to the participating organization using the program. Situations can be found readily where a program may have to originate at one place in the network coincident with the airing of a different program elsewhere on the network. Such situations are bound to occur due to scheduling problems within each participating organization. Thus, it would appear that at least two program channels that are independent of each other in their transmission are desirable.

A system with minimum operating costs has several aspects which must be considered. One type of economy is the use of FM radio relay to provide programs to all the stations rather than using more expensive wire lines. A second economy is to avoid the need for transmitter operators by use of transmitter remote control or by the location of the transmitters at sites of operating commercial stations that already are manned. A third economy can be through the use of a multiplexed subcarrier to provide a second program channel for relaying to certain areas or to special audiences. A fourth, and not so obvious economy of operation, would be by the use of equipment that is reliable and stable in operation, requiring a minimum of maintenance, upkeep, and attention. Other economies are possible but they should not be carried beyond the point where the program services will be adversely affected. A point can be reached for the present state of the art of broadcasting where further economies should not be considered.

FM Allocations Structure

The selection of a suitable FM channel (or frequency) for a particular city has to satisfy two basic considerations of the study. First, the proposed operation cannot cause objectionable interference to any other authorized FM station. This usually requires consideration of all FM stations within at least 200 miles of the proposed site. It also requires consideration of the television channels in the several areas, since use of certain FM frequencies can cause serious interference to television reception on some channels. Second, the proposed station must be received reliably and without interference at the next adjacent city in the relay chain to allow radio relaying of the programs throughout the network. The use of high effective radiated powers and antenna heights are proposed in order to provide large service areas and to improve the reliability of the radio relay network operation.

System Technical Details

The network would use conventional FM radio relay methods between the several stations. In order to secure reliable transmission over certain of the longer distances (75 miles or more), some diversity of the receiving antennas may be employed. Long time recordings of signal strength should be made to determine signal reliability before a final transmitter site has been chosen. This reception reliability is vital since several aspects of the system depend on it.

In addition to relaying the main channel programs, a second audio program channel could be relayed via a multiplexed subcarrier at about 67 kilocycles. This second audio program channel would be completely independent of the main program on the station, and would permit simultaneous two-way program relaying. Since actual program material would be transmitted, this subcarrier channel must have good response, high signal-to-noise ratio, and freedom from cross talk.

Control of the network technical operations could be handled remotely at several stations by means of an audio channel multiplexed on each of the stations at about 41 kilocycles. This channel would carry coded tones to convey the control information to the remote stations. Other coded tones would be used for the return of the remote transmitter information using this same subcarrier frequency multiplexed on the remote station.

When there would be only one remotely controlled station, the control channel could become a 50 to 5000 cps information channel in each direction, i.e. the full subcarrier response of the control station and the remote station would be utilized. There would be no problem of feedback since there would be no feedback connection between the received information and the outgoing control information. However, if a second remote station is to be controlled through the first remote station, subchannel feedback could occur unless the audio bandwidth is split to permit one-half the bandwidth to be relayed in the outgoing direction and the other half of the bandwidth to be the return relay to the control station. In this way, several stations could be controlled through each other using audio filters to prevent subchannel oscillation (feedback) and instability.

The remote control of the transmitters and the network switching would be broken down into four general functions, namely: audio switching control; transmitter function control; transmitter indicators; and supervisory indicators. These are each discussed in some detail to provide an understanding of each of the functions.

Fig. 1 is a block diagram layout for a typical control station for two remote stations. Additional remote stations would require additional receivers, diversity controls, frequency and modulation monitors and RF preamplifiers. Present FM Rules do not permit regular operation without a frequency and modulation monitor. Therefore, remote transmitter control is limited to those distances where adequate signal can be received for driving the frequency and modulation monitor. New models of RF preamplifiers are available which should provide ranges of 60-90 miles with good receiving locations.

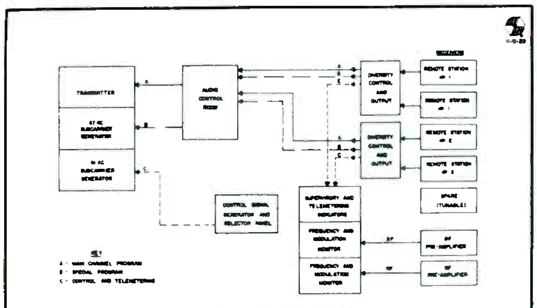

FIG. 1
BLOCK DIAGRAM FOR CONTROL STATION
EDUCATIONAL RADIO NETWORK
WGBH EDUCATIONAL FOUNDATION
CAMBRIDGE MASS

Fig. 2 is a similar block diagram for the facilities at a remote station. It is laid out for the operation discussed above for controlling a second remote station through this remote station. All control signals to be relayed through this station are routed through appropriate high or low pass filters to prevent oscillation or feedback on the control subcarrier channel.

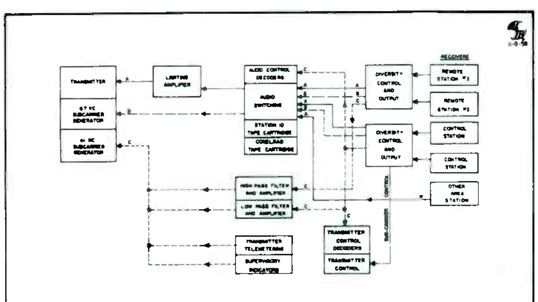

FIG 2
BLOCK DIAGRAM FOR REMOTE STATION
EDUCATIONAL RADIO NETWORK
WGBH EDUCATIONAL FOUNDATION
CAMBRIDGE, MASS

The remote audio switcher (see Fig.2) would enable programs to be routed through the network from any desired originating station. This might require some stations to have many different program sources including area stations not directly affiliated with the network. The switcher would also insert recorded station identifications from a continuous loop magnetic tape cartridge. The CONELRAD test and alerting announcements, where required, could be done with a second tape cartridge. A limiting amplifier would be used at the transmitter input to prevent main channel overmodulation. The audio switcher could, if desired, switch the undemodulated program subcarrier directly into the transmitter, when program subcarrier relay is used. This would help the maintenance of higher transmission quality when the subcarrier is relayed through several stations.

The transmitter function control would include: Transmitter On/Off; Plate On/Off; Plate Tune, and Power Output Raise/Lower. A fail-safe circuit would provide an automatic "Plate Off" control approximately 30 seconds after the control station subcarrier goes off the air, and a "Transmitter Off" 20 minutes later if the control station subcarrier has not returned to the air. Receivers and control equipment would remain "On" at all times. The tower lights would be automatic in operation from a PE cell. Protection for undervoltage and partial power failures should be incorporated within the transmitter.

The transmitter meter readings (Plate Voltage, Plate Current, and output) could be coded into amplitude or pulse time variations of a tone for transmission to the controlling station. A calibration voltage reference would be used to insure system accuracy. The meters would be sampled by properly coded interrogation signals from the control station.

Supervisory indicators could be relayed to the control point by means of coded tones. These would be selected just as the meter readings, except that the "reading" would be a short tone burst in response to the interrogation signal. One frequency could be a positive reply and a different frequency a negative reply. Continuous supervision indicators are not anticipated to be required. However, as an option, regular checking by means of a motor driven sampling switch could prove desirable. The frequency and modulation monitor would usually be operated at the control point from an "Off-Air" pickup using an RF preamplifier to drive the monitor. It is possible to remotely read the frequency deviation just as a transmitter meter, and the modulation directly "off the air" with a calibrated receiver. The Federal Communications Commission Rules do not permit this at the present time, however.

The transmitting antenna system would be six-bay conventional side mounted antennas fed with 1 5/8-inch diameter semiflexible coaxial transmission line. The antenna would have a gain of 7.5 to 8 db above a standard reference dipole to provide an effective radiated power of approximately 25 kilowatts when used with a five-kilowatt transmitter. The voltage standing wave ratio on the antenna and transmission line should be kept as low as possible so as to present the best possible load to the transmitter for the multiplexed transmissions. Antenna deicing equipment and weather protection for the transmission line would be provided where local conditions require their use. Certain installations might require different antenna and transmitter combinations in order to fit specific mounting, space or other limitations present at the site.

It is anticipated that the initial stages of network operation would omit the 67-kilocycle multiplexed subcarrier channel to expedite the start of the network. With only the main program channel and the 41-kilocycle multiplexed subcarrier for control, production type multiplex receivers could be used with a slight modification to improve the audio output circuits.

When the 67-kilocycle subcarrier is incorporated in the network operation, it might be desirable to consider a more complex receiver which would provide all of the outputs needed from a single receiver. These outputs would be: Program audio from the main carrier; control tones from the 41-kilocycle subcarrier; program audio from the 67-kilocycle subcarrier; and 67-kilocycle undemodulated subcarrier to insert directly into the transmitter for subsequent subcarrier relaying. Such a receiver is not known to be currently available except on a custom built basis but could be produced in the future if subcarrier relaying is done extensively.

The receiving antennas would be ruggedized high gain type yagis for durability and good directivity. Other types such as rhombics and broadside arrays could be used where special needs arise. All receiving antennas would use coaxial transmission line to reduce undesired pickup.

Proposed Equipment Specifications

The following specifications for equipment, conceived as suitable for the foregoing purposes, are detailed with due attention to certain system specifications which are made a part of the requirements outlined.

Transmitters

The transmitters shall be of 5 or 10 kilowatt rating equipped with two subcarrier generators to provide a total of three audio channels. The transmitter used shall conform to all applicable Federal Communications Commission Rules and Standards as well as EIA TR-107 Standards except where detailed herein. Normal operation of the transmitter will require the rated power output into a 50.0 ohm load.

The audio channels shall be designated as follows: Main channel - "A"; subchannel modulating a 67-kilocycle subcarrier - "B"; subchannel modulating a 41-kilocycle subcarrier - "C".

Channel "A" shall have the following characteristics when the transmitter output is sampled by a demodulator of suitable quality. Seventy-five microsecond de-emphasis shall be incorporated in the demodulator.

Frequency response:	± .5 db, 50-15000 cps
	±.25 db, 100-12000 cps
Harmonic distortion:	1% maximum for frequencies 50-15000 cps
	.5% maximum for frequencies 100-7500 cps
FM noise level:	70 db below 75 kc deviation
AM noise level:	55 db below 100% amplitude modulation
Crosstalk when Channels "B" and "C" are each modulated 7.5 kc:	60 db below 60 kc deviation

Channel "B", the 67-kilocycle subcarrier channel, shall have the following characteristics when the transmitter output is sampled by a suitable demodulator. Seventy-five microsecond de-emphasis shall be incorporated in the demodulator. One-hundred per cent modulation of the 67-kilocycle subcarrier shall be 7.5 kilocycle deviation.

Frequency response:	±1.5 db, 50-5000 cps
Harmonic distortion:	3% maximum for frequencies 50-5000 cps

FM noise level: 55 db below 7.5 kc deviation

AM noise level: 50 db below 100% amplitude modulation

Crosstalk when Channel "A" is modulated to 60 kc deviation and Channel "C" is modulated to 7.5 kc deviation: 45 db below 7.5 kc deviation

The transmitter shall incorporate adequate protection circuitry to prevent damage or component failure due to low voltage or loss of one or more phases of primary input power. The transmitter shall be temperature stable to allow unattended operation with ambient temperatures from 40 to 120 degrees Fahrenheit.

Transmitting Antenna System

The transmitting antennas shall be six-bay conventional type having 7.5 to 8.0 db gain over a standard tuned dipole. The transmitting antenna shall conform to EIA Standards TR-117 except where detailed herein. The input power rating shall exceed 10 kilowatts for all frequencies 88 to 108 megacycles. The horizontal pattern circularity shall be ±1 db in free space when side mounted on a tower having a maximum face dimension of 18 inches. Input impedance shall be 50 ohms. Input VSWR shall be 1.1 to 1 or better for all frequencies ±100 kilocycles of the transmitter carrier. Antenna deicers will be required at certain installations. The deicers shall be automatic in operation from a thermostatically controlled switch near the antenna.

The transmission line shall be 1 5/8-inch semiflexible 50-ohm coaxial cable with a habelin or vinyl plastic outer covering. The transmission line shall be attached to the tower with stainless steel Wraplock. The transmission line shall have an input VSWR of 1.08 to 1 or better for continuous lengths up to 250 feet when terminated by a 50-ohm resistance load. Measurements shall include the end adaptors for connection to the transmitter and antenna. The horizontal run of transmission line shall have adequate support and protection from hazards such as falling ice.

Remote Control

Outgoing control signals shall originate at the control station from stable tone generators using frequencies selected in the range of 50 to 500 cps at such intervals as to avoid harmonic and intermodulation problems between the several tones. Each control function shall be actuated by the proper tone sequential or tone group sequential code. The tone code combinations shall be selected by the actuation of push buttons, each push button representing one control function. A suitable number of tone generators and coding combinations shall be provided to allow for at least 300 different control functions as the network continues to grow. Approximately 30 control functions shall be possible for each remotely controlled station.

The control signal decoders at the remote transmitter shall be insensitive to voice communication in the frequency range of 300 to 1000 cps on the same circuit. The outgoing control signal Channel "C" shall be band limited by a 1000-cycle low-pass filter giving 40 db attenuation at 1500 cps for reception by any remote station beyond the first.

The outgoing control signals shall actuate the following functions at each remote station: One of six different audio inputs for the main program Channel "A"; one of six different audio inputs for the program Subchannel "B"; Station Identification* (Channel "A" only); CONELRAD Test or Alert*; Transmitter "On"; Transmitter "Off"; Plate "On"; Plate "Off"; Plate Tune Right; Plate Tune Left; Power Raise; Power Lower; Calibrate; PA Volts; PA Current; Output; Tower Lights Check; Temperature Check; Security Check; and such others as the individual installation may require. Some stations will have only the audio switching controls when they are located at a manned transmitter site where the routine transmitter operations can be handled by the operator on duty.

The audio program switcher at the remote station shall have the required control signal decoders to actuate suitable type relays with DPDT wiping contacts equipped with locking circuits. All signal decoders and any tone relays (if used) shall be plug-in type and sealed from dust and dirt. Hermetic sealing is desirable. The locking circuits shall release upon closure of another audio relay or upon station identification to prevent two programs on the same channel at once. A limiting amplifier shall be supplied to prevent overmodulation of the main program Channel "A".

Upon reception of the proper control signal, station identification shall be made from a tape recorder with automatic cuing after usage. The CONELRAD alerts

---

*Station Identification and CONELRAD functions may use the same control signals for all stations.

and tests will be made with a similar unit except that additional means shall be provided for cycling the transmitter plate "On" and "Off" if required by the CONELRAD procedures.

Transmitter control shall be accomplished by suitable decoders terminating in appropriate relays for the controlled function. In the case of the "plate tuning" and "power output", a reversible motor and gear reduction drive coupled to the tuning shaft shall be the controlled function. Limit switches or other suitable means shall be incorporated to prevent damage at the extremes of the control range. Tuning by incremental steps is acceptable.

A "fail-safe" circuit shall be incorporated in the transmitter control to turn the transmitter plate "Off" after loss of the control station's sub-carrier for 30 seconds. The transmitter shall then automatically turn "Off" 20 minutes later if the control station sub-carrier has not returned to the air.

Return control signals that originate at the remote stations shall be from stable tone generators using frequencies selected in the range of 1500 to 5000 cps at such intervals that will prevent harmonic and intermodulation problems between the several tones. One tone shall be used for all telemetering from the several remote stations. Two other tones shall be used to provide the 3 to 5 second supervisory indications when interrogated by the control station. For normal operation, the Subchannel "C" shall have no modulation on it except when the control station is checking or changing the operation of the network. The return control signal on Channel "C" shall be band limited by a 1500-cycle high-pass filter providing 40 db attenuation at 1000 cps for reception from any remote station beyond the first.

Telemetering shall be by amplitude variation of the telemetering tone. Tone pulse duration may be substituted for amplitude variation if a zero to full scale reading can be achieved with ±5 per cent accuracy in less than two seconds time. The complete telemetering system including the readout meter must satisfy Federal Communications Commission requirements for accuracy including those given in Sections 3.320 and 3.39(d)(2) of the Rules and Standards.

The supervisory indications shall consist of positive or negative answers to selected operations that must be monitored at regular intervals. Any failure of the supervisory indicator circuit shall result in absence of an answer. A false answer, whether positive or negative shall not result from any failure of the supervisory indicator circuit.

The required frequency and modulation monitors for the remotely controlled transmitters shall be located at the control point if possible. Suitable preamplifiers and receiving antennas shall be incorporated to permit "Off-Air" reception of the remote station by the monitor. The preamplifier and antennas shall be separate from all other receiving equipment at the control point.

Receiving Equipment

The receiving antennas shall be 10-element yagi type, ruggedized for durable service under adverse weather conditions. The driven elements of the antenna shall be folded dipole design with impedance matching to an unbalanced 75-ohm Type N coaxial fitting. The inside of the matching section shall be impregnated and filled with insulating compound to prevent leakage or condensation of moisture therein. Mechanical dampening shall be used on all elements of the antenna. Construction shall be of high strength aluminum alloy with all hardware of stainless steel. Welded and lock-bolted construction shall be used throughout. Undesired signal rejection relative to the maximum lobe, oriented at zero degrees, shall exceed 26 db for all azimuth angles between 70 and 290 degrees in the horizontal plane. Forward gain over a tuned reference dipole shall exceed 10 db for the frequency range 88 to 96 megacycles. It is anticipated that certain locations may require special additional features such as deicing equipment, and antenna stacking of both a vertical and horizontal nature. Accessories shall be available for such special arrangements. Support masts and extension arm hardware shall be of adequate mechanical design.

The receivers shall be fixed tuned, crystal controlled type with the following characteristics:

Tuning range: 88-108 mc with proper crystals

Sensitivity: 10 microvolt signal across receiver input shall provide at least 40 db signal-to-noise ratio on all outputs

Input impedance: 75 ohms unbalanced, Type N coaxial fitting

Undesired channel rejection: (Main channel d/u input ratio for 50 db d/u main channel output ratio)	0 kc, +4 db or better 200 kc, -5 db or better 400 kc, -50 db or better 600 kc and over, -60 db or better

Crosstalk: The signal-to-noise ratios given under the receiver outputs shall apply also to crosstalk from all other channels

Simultaneous receiver outputs:

Main Channel "A"
50-15000 cps, ±1.5 db response
65 db signal-to-noise ratio
1% harmonic distortion, 50-15000 cps, 75 kc deviation
0 dbm, 600 ohm output

Subchannel "B"
50-12000 cps, ±1.5 db response
55 db signal-to-noise ratio
1.5% harmonic distortion, 50-12000 cps, 7.5 kc deviation
0 dbm, 600 ohm output

Subcarrier "B"
55-75 kc ±1.5 db response
60 db desired-undesired signal ratio
±1% output level regulation
0-2 volts, 75 ohm coaxial output

Subchannel "C"
50-5000 cps, ±3 db response
45 db signal-to-noise ratio
5% harmonic distortion, 50-5000 cps, 7.5 kc deviation
0-2 volts high impedance output

Limiter voltage: High impedance DC output

In the initial stages of the network operation, it is anticipated that the operating requirements may not justify all the previously mentioned requirements for receivers. Upon full implementation of the network, such receivers will be needed for maximum reliability and flexibility. For the start of the network operation, high quality production multiplex receivers with slight output modification would appear to be satisfactory.

THE USE OF PHASE EQUALIZERS TO IMPROVE THE TRANSIENT RESPONSE OF A TELEVISION TRANSMITTING SYSTEM

J. K. MacDonald
Plant Department
Canadian Broadcasting Corp.
Montreal, Quebec

Since the inception of the present system of generation and transmission of video information, it has been realized that amplitude distortions and phase distortions both contribute to degradation of the quality of the picture. A great deal of work has been done, particularly in Europe, on the use of transient waveforms, for the direct observation of the effects of distortions of the amplitude vs frequency and phase vs frequency characteristic.

With this in mind it was decided to apply the technique of transient testing in conjunction with the use of phase equalizers to that part of the overall video system which is inherently one of the greatest offenders in producing phase distortions, namely the system comprising the transmitter and receiver, the object being to improve the fidelity of picture reproduction.

The transient waveform used was the sine-squared pulse. It has several desirable characteristics for this purpose which are well described in the literature.[2,3,4,5,6].

The phase equalizers used were those commercially available for the envelope delay correction of color T.V. transmitters and comprise the following elements:

1. A variable Low Frequency Phase Equalizer.
2. A High Frequency Fixed Delay.
3. A Variable High Frequency Equalizer.
4. A Fixed Receiver Equalizer.
5. A Fixed Notch Equalizer.

Theory

Refer to Figure 1. This figure represents the amplitude vs frequency characteristics of an ideal receiver (solid) and an ideal transmitter (dotted).

For this application of transient testing technique the following important observations may be made from an examination of Figure 1.

1. It is necessary to consider the transmitter and receiver as a system in order to maintain proper amplitude balance of upper and lower sidebands.

2. It is possible to check the performance of the receiver using a double sideband test transmitter.

In a combination of vestigial sideband transmitter and receiver certain phase distortions are inherent as follows:

1. Low frequency distortions due to the vestigial sideband reception in the receiver.

2. High frequency distortions due to the cut-off characteristic of the receiver.

3. Similar Low and High frequency distortions due to the transmitter cut-off characteristics. These distortions in terms of relative phase delay (Ø/W) are of the form shown in Figure 2.

Distortions of this type cause characteristic distortions of the shape of the sine-squared pulse as shown in Figure 3. An examination of this figure shows that one of the advantages of the use of the sine-squared pulse is that it is possible to separate the effects of high frequency distortion and low frequency distortion by the judicious use of the 0.125 micro-second and the 0.25 micro-second half amplitude duration pulses. These are referred to as the T and 2T pulses respectively.

A. Shows the effect of low frequency phase distortions on the 2T pulse.

B. Shows the effect of low and high frequency phase distortions on the T pulse.

C. Shows the effect of passing the T pulse through an ideal low pass filter.

Description of Tests

It was decided before the tests were begun to adjust the phase equalizers for best reproduction of the 2T pulse after calibration of the station demodulator, i.e. to compensate for receiver low frequency distortions in addition to the transmitter low frequency distortions. Kell and Fredendall[7] have shown that receiver distortions due to vestigial sideband reception are quite similar for a number of different receivers. It was therefore felt that compensation for the characteristic of the station demodulator would afford some degree of compensation for most well adjusted home receivers.

Since the demodulator is to be used as part of the test equipment, it is necessary to calibrate its performance and to establish limits for its response to the 2T and T pulses. Tentative limits and their derivation will be given later in this article.

The following is a list of the required test equipment.

1. Sine-Squared Pulse Generator.
2. Test Transmitter.
3. Band Pass R.F. Amplifier if necessary to raise the level or to eliminate spurious radiations in the Test Transmitter.
4. Station Demodulator.
5. Oscilloscope.
6. Sine-Squared Pulse Graticule.
7. Envelope Detector.

The characteristics of the test transmitter in item 2 above should be as follows:

a) Freedom from low frequency phase distortion (achieved by double sideband operation).
b) Freedom from high frequency phase distortions within the pass band of the station demodulator (achieved by suitable broad-banding of the test transmitter).
c) Good performance in terms of standard T.V. transmitter parameters, e.g. amplitude vs frequency response, linearity, etc.

The test procedure may be summarized as follows:

1. Check on the steady state and transient performance of the test transmitter using a diode detector.
2. Adjustment of the R.F. Pass band of the demodulator to a characteristic as near the ideal receiver characteristic as possible but with the response flat within 2db to 5.5 mc/s above visual carrier.
3. Check of the transient performance of the demodulator.
4. Steady state alignment of the transmitter.
5. Transient testing of the transmitter using the 2T pulse and adjusting the Low Frequency Phase equalizer for best results. (The demodulator is used as the detector in this and all future tests.)
6. Check on high frequency phase distortions of the transmitter using the T pulse and adjusting the High Frequency equalizer for best results.
7. Observation of the effect of the demodulator aural trap(s) on the T pulse response.
8. Observation of the effect of the Receiver equalizer using the T pulse.

Figures 4 to 15 are photographs showing these effects as follows:

Figure 4. 2T pulse through the test transmitter and demodulator.

Figure 5. T pulse through the test transmitter and demodulator.

Figure 6. 2T pulse through the test transmitter and demodulator no phase compensation.

Figure 7. 2T pulse through transmitter and demodulator after adjustment of the low frequency phase equalizer.

Figure 8. T pulse through transmitter and demodulator.

Figure 9. T pulse through transmitter and demodulator after adjustment of the high frequency phase equalizer.

Figure 10. T pulse through the transmitter and demodulator after switching in the demodulator aural traps.

Figure 11. T pulse through transmitter and demodulator with Receiver equalizer in circuit.

Figure 12. Transmitted monoscope before adjustment of phase equalizers.

Figure 13. Transmitted monoscope after adjustment of all phase equalizers.

Figure 14. Monoscope at input to transmitter modulator after adjustment of all phase equalizers.

A block diagram of the test equipment set up for calibration of the demodulator is shown in Figure 15.

Tentative Demodulator Specification

Tentative limits of the performance of the demodulator used in this application have been established.

Kell and Fredendall[7] have given tentative limits for the response of a standard receiver to a square wave.

These limits were modified slightly and the limits used appear in Figure 16.

The equivalent 2T and filtered T pulse response corresponding to the possible extreme conditions of Figure 17 were derived using the Time-Series Method of Calculation of Lewis[8] and appear in Figures 17 and 18. Figure 17 shows the equivalent 2T pulse. The envelope shows tentative limits for the demodulator response and corresponds to a rating factor K = 4%.[4,5] As more experience is gained these limits may be modified if necessary. Figure 18 shows the equivalent T pulse. It is anticipated that the T pulse response including the overshoots will not greatly exceed the limits K = 4%. Additionally limits for the T pulse width are shown.

Conclusion

It has been found possible to effect a considerable improvement in the picture quality of a typical T.V. transmitter system using transient testing techniques and phase equalization. The method has been found to be straightforward and rules for its application may be set up such that particularly specialized knowledge is not required.

Bibliography

1. Kell R. D. & Fredendall G. L. - Selective Sideband Transmission in Television - RCA Review, April 1940, p. 425.

2. Hunt G. J. & Elliott E. W. - An Introduction to the Sine-Squared Pulse - Journal of the Television Society, Volume 7, No. 2, Page 49.

3. Kennedy Ralph C. - Sine-Squared Pulses Test Color - T.V. Systems - Electronics, December 1954, Page 138.

4. Rendall A. R. A. - Waveform Testing Methods for Television Links - Electronic & Radio Engineer, December 1957, Page 451.

5. Lewis N. W. - Waveform Responses of Television Links - Proceedings I.E.E., Part III Volume 101, July 1954, Page 258.

6. Macdiarmid I. F. - Proceedings I.E.E., Part IIIA Volume 99, 1952, Page 436.

7. Kell R. D. & Fredendall G. L. - Standardization of the Transient Response of Television Transmitters - RCA Review, March 1949, Page 17.

8. Lewis N. W. - Waveform Computations by the Time-Series Method - Proceedings of the I.E.E. Part III, Volume 99, 1952, Page 294.

Acknowledgements

The author wishes to express his thanks to Messrs. W. G. Richardson, Director of Engineering, and J. E. Hayes, Chief Engineer, for permission to publish this paper and to Mr. J. Carlisle, Plant Engineer under whose general supervision the work was carried out. Special thanks are due to Messrs. A. Ste-Marie, E. G. Elliott, D. MacRae and the staff of CBLT Transmitter for their invaluable assistance during the tests.

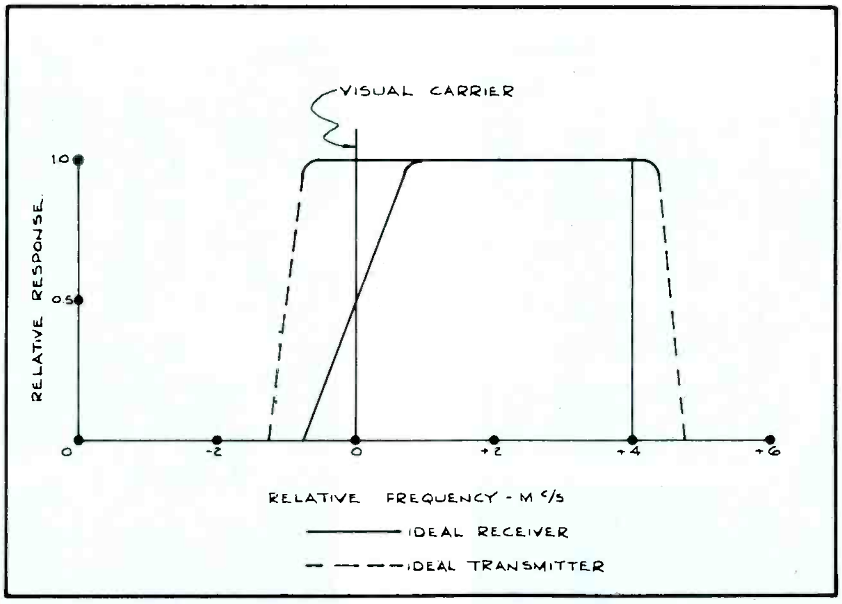

Fig. 1

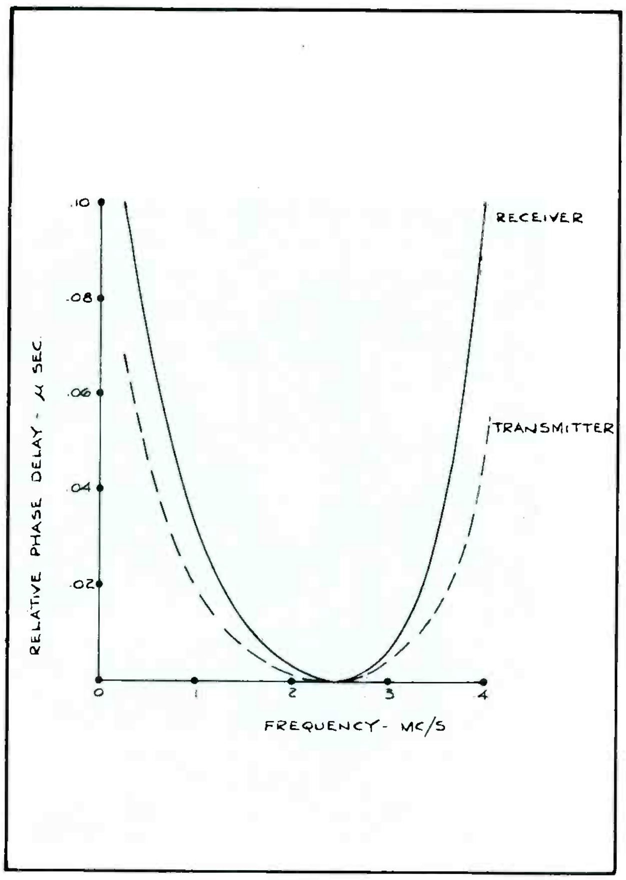

Fig. 2

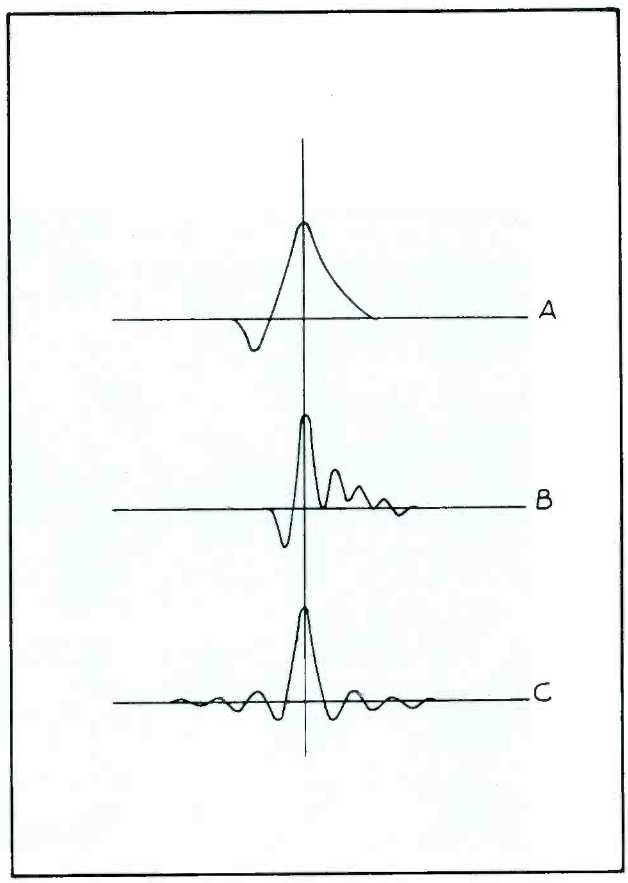

Fig. 3

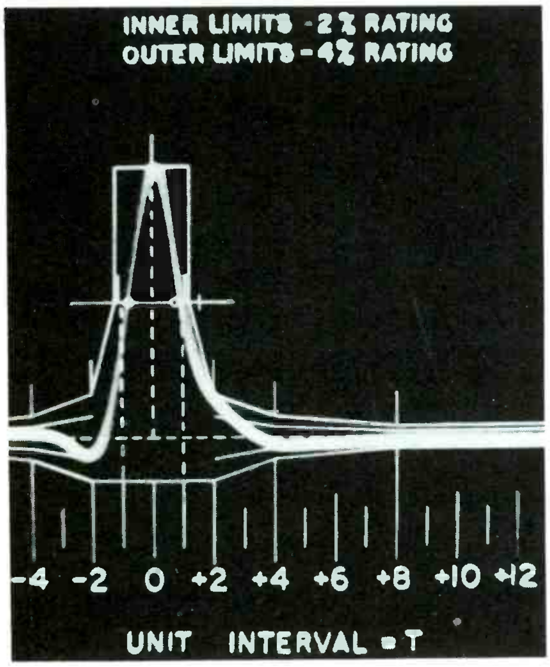

Fig. 4 — T-0.125 microsecond.

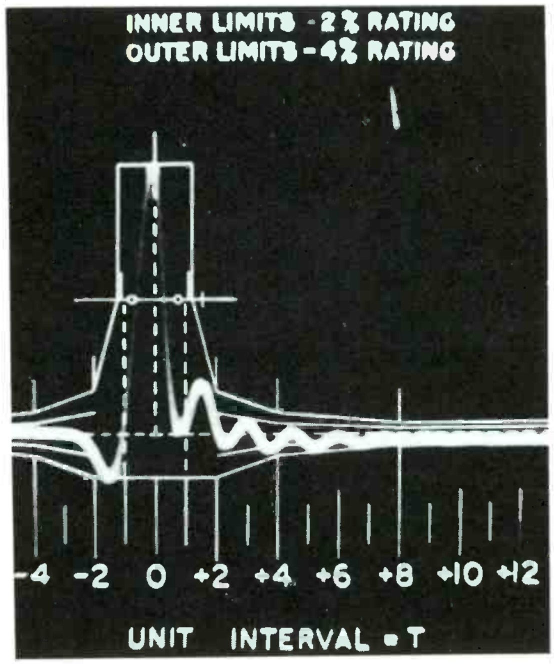

Fig. 5 — T-0.125 microsecond.

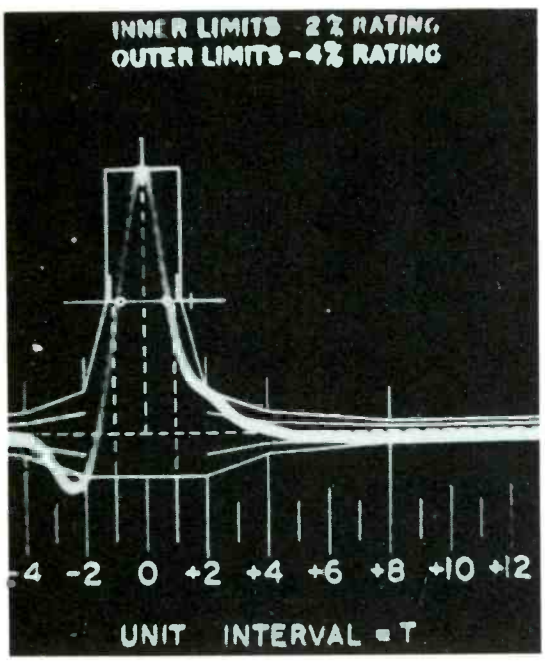

Fig. 6 — T-0.125 microsecond.

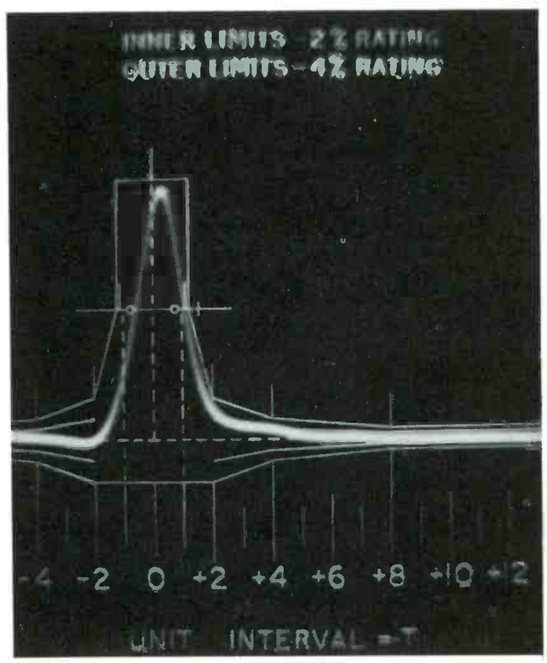

Fig. 7 — T-0.125 microsecond.

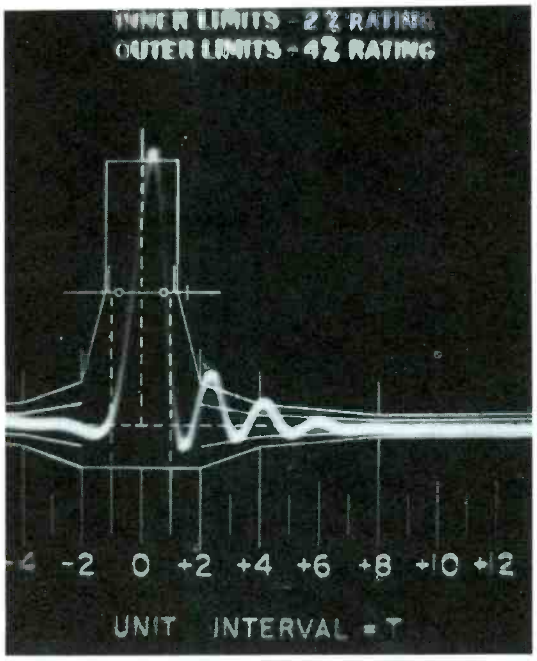

Fig. 8 — T-0.125 microsecond.

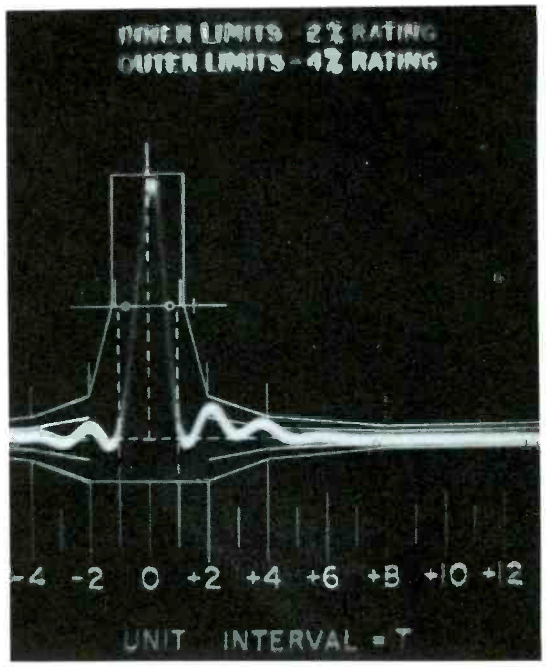

Fig. 9 — T-0.125 microsecond.

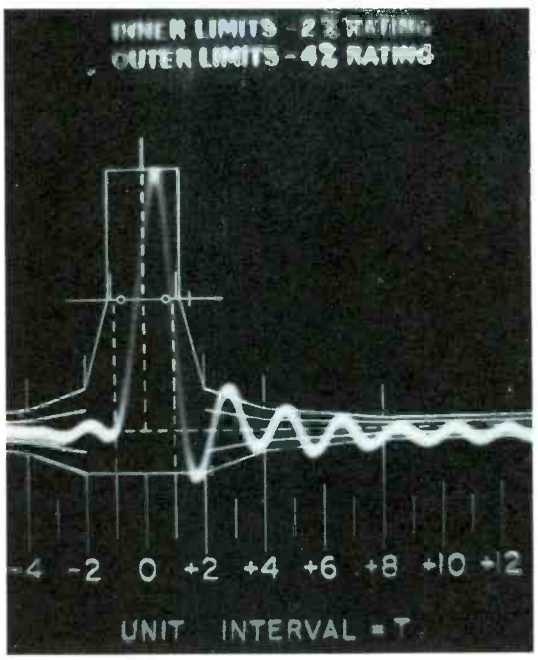

Fig. 10 — T-0.125 microsecond.

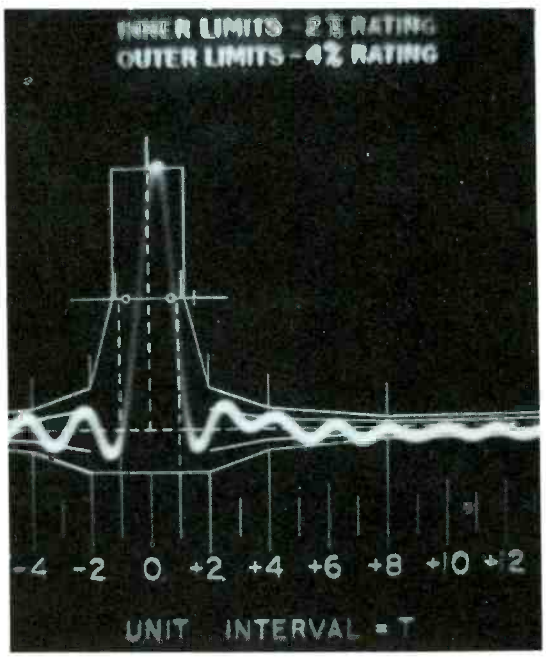

Fig. 11 — T-0.125 microsecond.

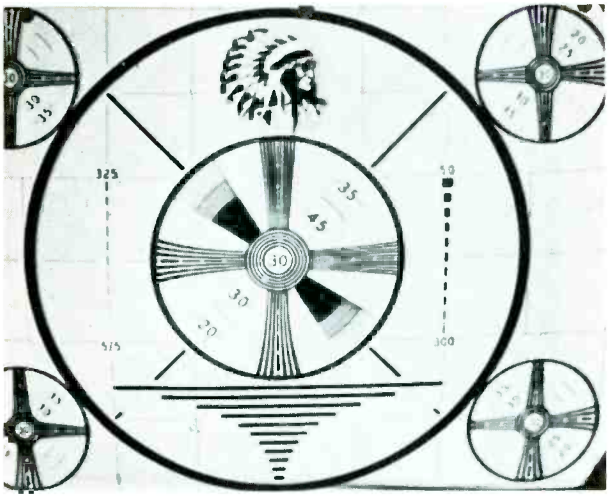

Fig. 12

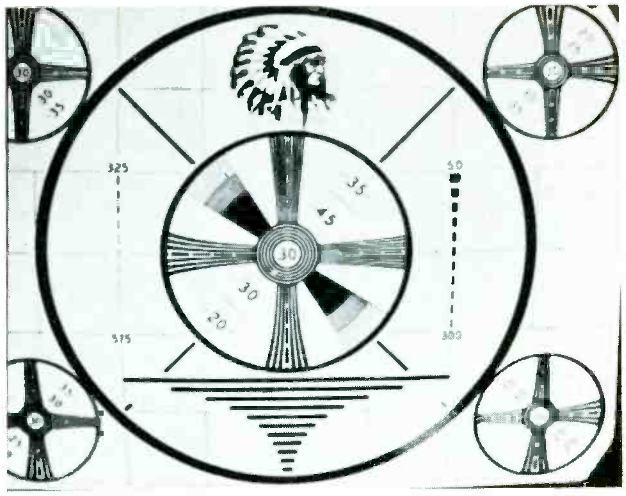

Fig. 13

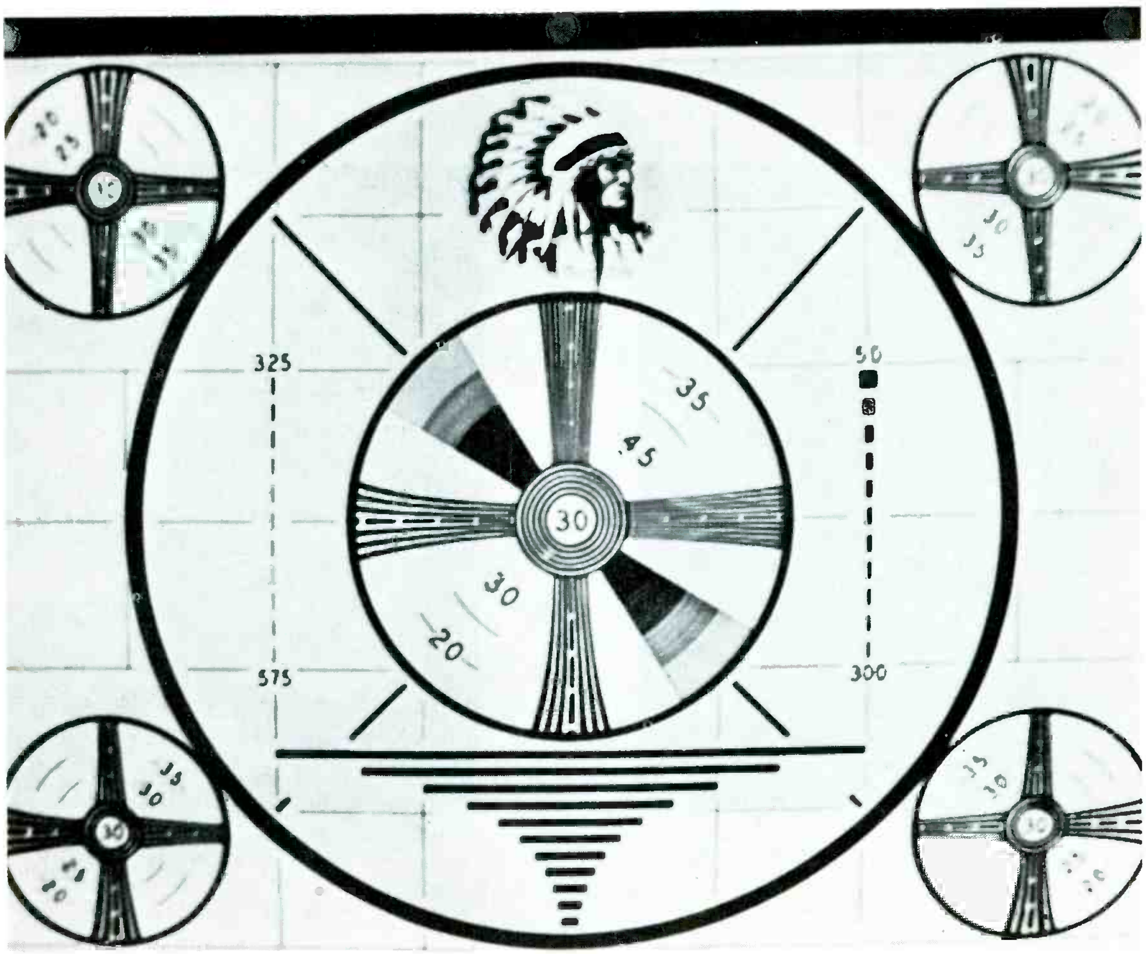

Fig. 14

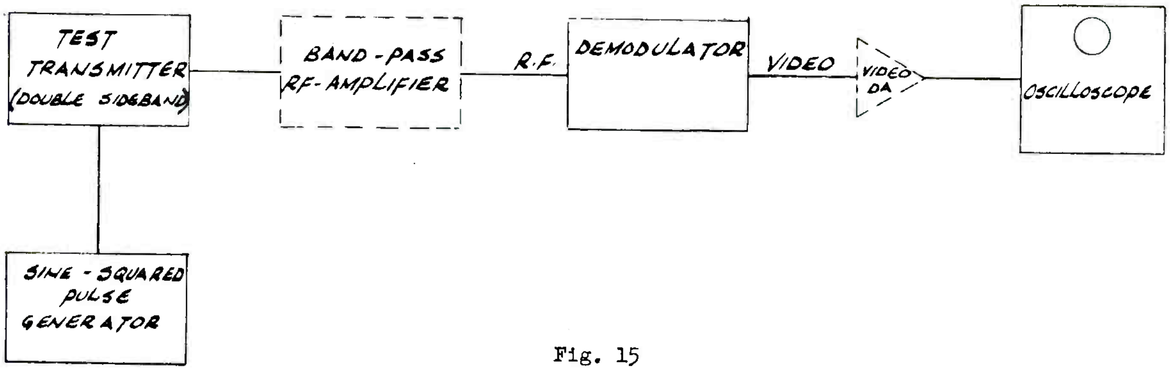
TEST
TRANSMITTER
(DOUBLE SIDEBAND)
BAND-PASS
RF-AMPLIFIER
R.F.
DEMODULATOR
VIDEO
VIDEO
DA
OSCILLOSCOPE
SINE-SQUARED
PULSE
GENERATOR

Fig. 15

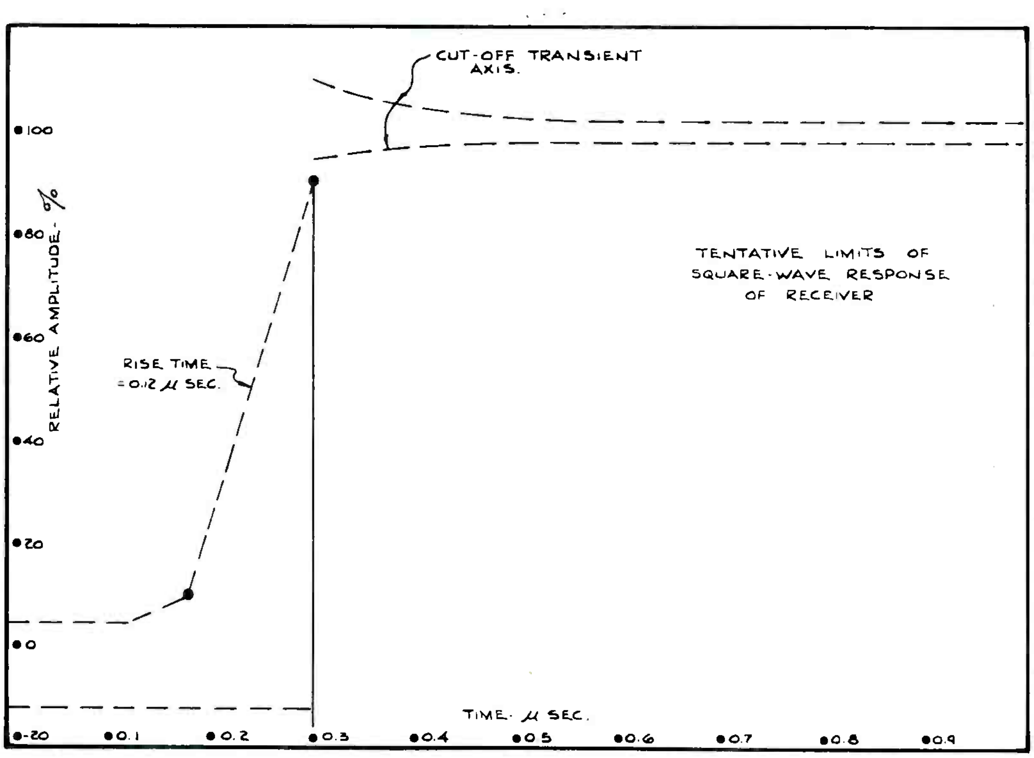
CUT-OFF TRANSIENT
AXIS.
100
80
60
40
20
0
RELATIVE AMPLITUDE - %
TENTATIVE LIMITS OF
SQUARE-WAVE RESPONSE
OF RECEIVER
RISE TIME
=0.12 μ SEC.
TIME- μ SEC.
-20
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

Fig. 16

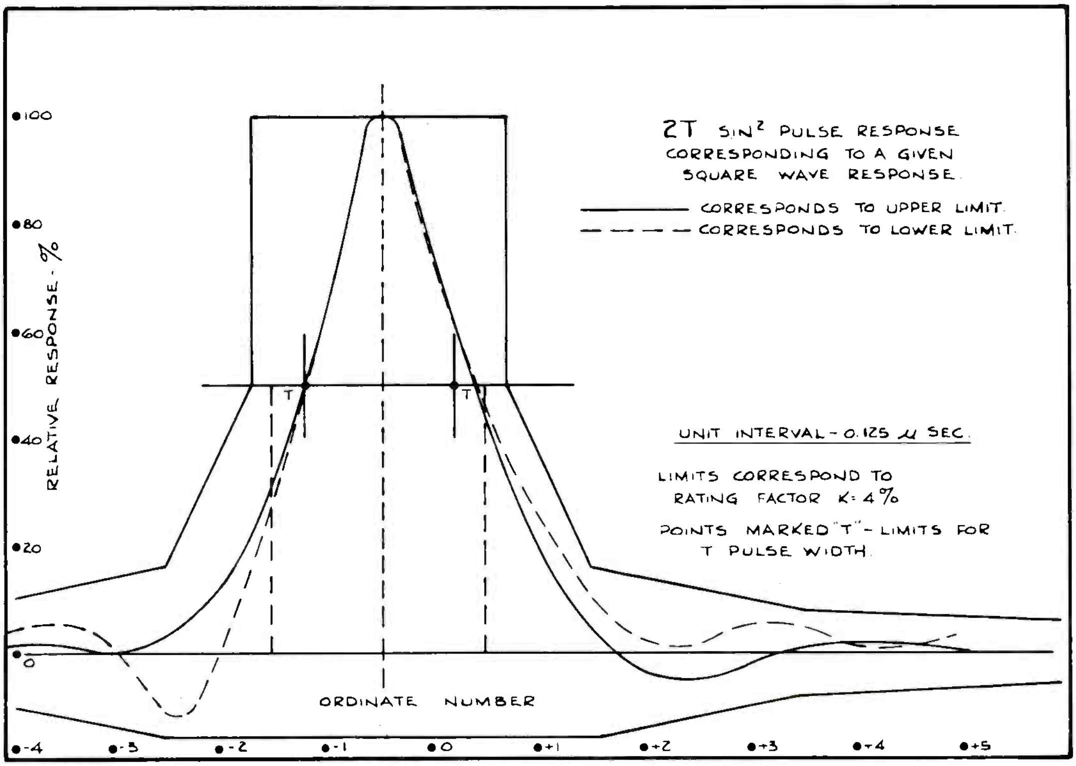
2T SIN² PULSE RESPONSE CORRESPONDING TO A GIVEN SQUARE WAVE RESPONSE.
CORRESPONDS TO UPPER LIMIT.
CORRESPONDS TO LOWER LIMIT.
UNIT INTERVAL - 0.125 μ SEC.
LIMITS CORRESPOND TO RATING FACTOR K= 4%
POINTS MARKED "T" - LIMITS FOR T PULSE WIDTH.
RELATIVE RESPONSE - %
ORDINATE NUMBER
100
80
60
40
20
0
T
T
-4
-3
-2
-1
0
+1
+2
+3
+4
+5

Fig. 17

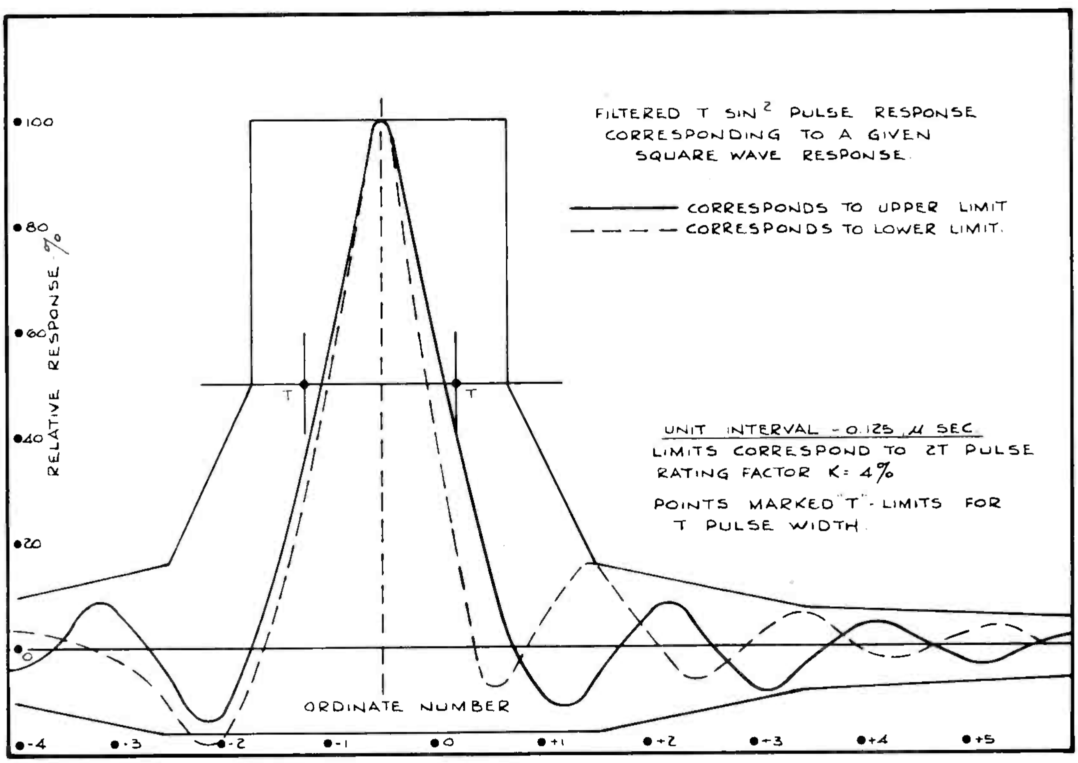
FILTERED T SIN² PULSE RESPONSE CORRESPONDING TO A GIVEN SQUARE WAVE RESPONSE.
CORRESPONDS TO UPPER LIMIT
CORRESPONDS TO LOWER LIMIT.
UNIT INTERVAL - 0.125 μ SEC
LIMITS CORRESPOND TO 2T PULSE RATING FACTOR K= 4%
POINTS MARKED "T" - LIMITS FOR T PULSE WIDTH.
RELATIVE RESPONSE %
ORDINATE NUMBER
100
80
60
40
20
0
T
T
-4
-3
-2
-1
0
+1
+2
+3
+4
+5

Fig. 18

A DESCRIPTION OF THE TUNNEL DIODE
AND ITS APPLICATIONS

John W. Wentworth*
Broadcast and Television Division
Radio Corporation of America

Introduction

The tunnel diode is one of the most significant solid-state electronic devices to emerge from the research laboratories since the transistor.

It is very simple in construction, consisting only of a p-n semiconductor junction with ohmic contacts to the p and n materials.

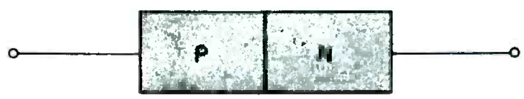

Figure 1

It differs from a normal junction diode, however, in that the semiconductor material is much more heavily "doped" with impurity atoms, resulting in a very significant change in the diode behavior. In a normal diode, current tends to flow rather freely if the diode is biased in the forward direction (that is, with the p-type material positive), but the current flow is almost negligible if the p type material is made negative relative to the n type material (provided the reverse bias voltage is appreciably less than the so-called "breakdown" potential). Thus, the characteristic curve of a conventional diode has this familiar shape.

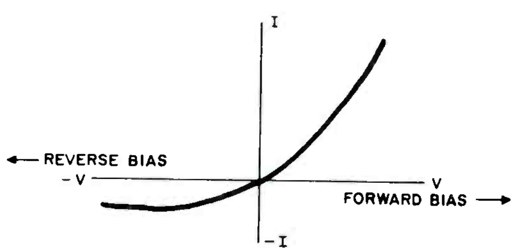

Figure 2

The tunnel diode, on the other hand, conducts very freely in the reverse direction and has an N-shaped characteristic in the forward direction. In a portion of the forward-bias region (from A to B on the diagram), the tunnel diode behaves like a negative resistance; it is this fact which accounts for most of its useful properties.

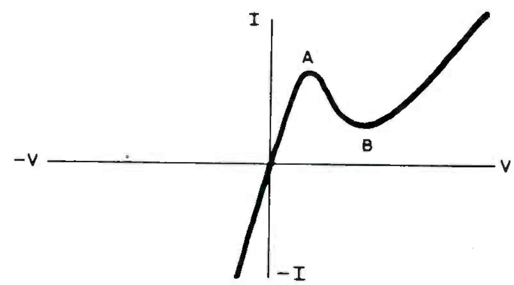

Figure 3

To explain how a tunnel diode operates, it should be helpful to review briefly some of the highlights of semiconductor physics. The treatment here is non-mathematical and somewhat superficial, but it should be adequate to give the reader at least a qualitative understanding of tunnel diode behavior.

Conduction in Metals

The conduction of electricity in any solid material depends upon the presence of electrons which are free to move within the molecular structure of the material. According to the quantum theory, the electrons in a given atom can exist only at discrete energy levels, and can jump from one energy level to another only by absorbing or radiating at least one quantum of energy. (The energy levels correspond roughly to the "orbits" assumed in the older versions of the atomic theory.) In a metallic conductor, the available energy levels are very close to each other, and the agitation of a solid crystal by thermal energy at all normal temperatures is sufficient to excite substantial numbers of electrons to energy levels slightly above the average level of the outermost "shell" of electrons. An electron energy diagram is frequently used to represent these conditions.

---

*This paper first appeared in Field Engineers' Electronics Digest (RCA) Vol. 8, No. 1, pp. 2-11, 1960.

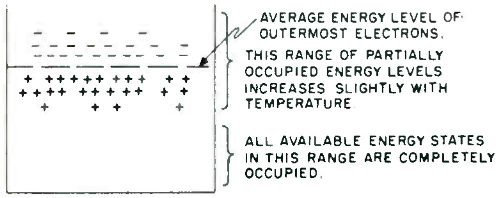

Figure 4

For every electron which has been "excited" by thermal energy so as to occupy an energy level above the dotted line in the diagram above (symbolized by minus signs), there is a corresponding "vacancy" or unoccupied energy state somewhere below the dotted line (symbolized by plus signs).

An electron at any of the partially-occupied energy levels is able to move rather freely through the crystal structure provided there are "vacancies" at the electron's energy level in neighboring atoms. Normally, the "shuttling" of electrons from one atom to another occurs in random fashion with no net movement of electric charge in any direction. Under the influence of an electric field, however, there is a net movement in the direction of the applied potential, constituting an electric current. Most metals are good conductors because their substantial range of partially-occupied energy states yields a good supply of so-called "free electrons" and of vacancies into which such electrons can move.

Electrical Properties and Semi-Conductors

In a highly-purified crystal of semiconductor material, the electrons in the outermost "shell" of each atom (that is, the valence electrons) are utilized to form the bonds between adjacent atoms. In a germanium crystal, for example, the four electrons in the outermost shell are used to form the inter-atom bonds symbolized by short lines in the 2-dimensional sketch shown in Fig. 5 (actually, the atoms form a 3-dimensional crystal lattice with double bonds, since each of the neighboring atoms contribute one bonding electron to hold the atom in the center).

To break loose from its bonding role and become free to move through the crystal structure, an electron must absorb a very substantial amount of energy. The energy level diagram for a semiconductor thus includes a substantial

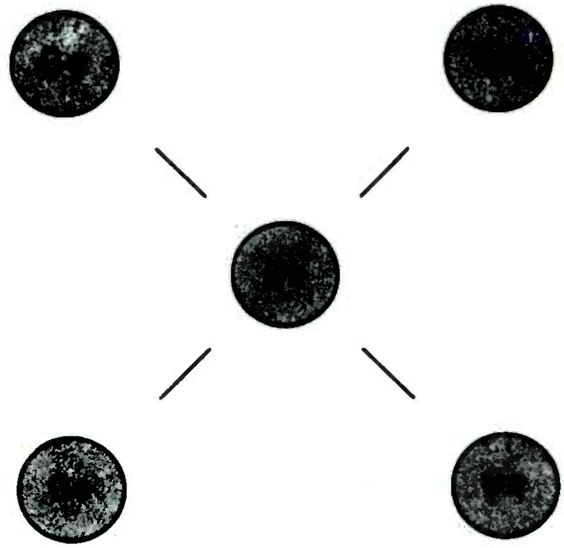

Figure 5

"forbidden region" within which no electron can exist. The available energy states above and below this forbidden region, are usually called the "conduction band" and "valence band," as shown in the diagram below.

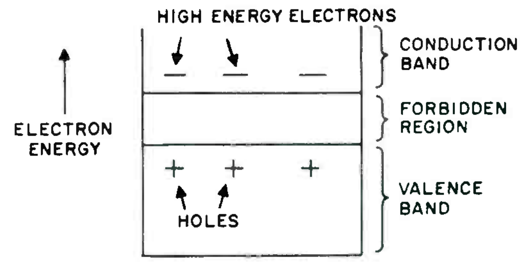

Figure 6

As a result of thermal agitation of the crystal, a few electrons acquire enough energy to hurdle the forbidden region, leaving "holes" or vacant energy states in the top of the valence band. The exact number of such free electrons per unit volume is a function of temperature, increasing at higher temperatures. Because there are relatively few "charge carriers" (free electrons or vacancies), a pure semiconductor is a very poor conductor of electricity. In fact, most insulating materials have the same type of energy diagram with such a wide "forbidden region" that virtually no electrons are able to absorb enough energy to hurdle the barrier at normal temperatures.

"Doping" of Semi-Conductors

For practical semiconductor devices, the pure semiconductor material is "doped" by the controlled addition of specific impurities to modify the electrical properties. So-called n-type germanium, for example, is formed by adding an impurity such as arsenic with five rather than four electrons in the outer "shell." In moderate quantities, atoms of the impurity material replace some of the germanium atoms in the crystal structure, and each "donates" one surplus electron not needed for bonding, as shown by the sketch below.

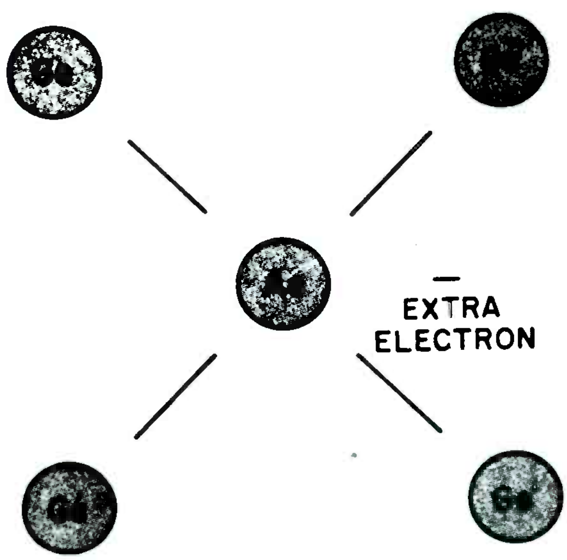

Figure 7

The extra electrons are readily excited to energy levels falling within the conduction band of the germanium crystal, so the energy diagram of an arsenic-doped germanium crystal is modified as shown below.

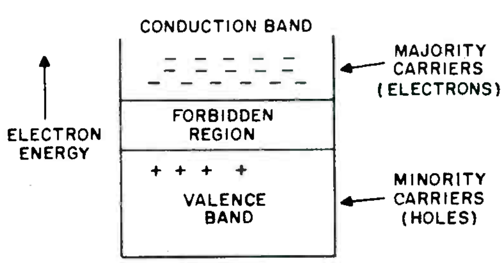

Figure 8

As in the case of the pure germanium crystal, a few of the electrons normally associated with the bonding function are able to absorb enough energy to break loose and become charge carriers, leaving behind a corresponding number of vacancies or "holes" in the valence band. Most of the electrons in the conduction band, however, are "donated" by the impurity atoms. Both the electrons and the "holes" are available as charge carriers. Electrons move directly, usually in short hops between neighboring atoms. "holes" only appear to move; when an electron from a given atom moves over to occupy a vacant energy state in an adjacent atom, it leaves behind a vacancy that has apparently moved in the opposite direction. Thus "holes" behave as if they were positively-charged carriers of electricity, even though it is always electrons that do the actual moving. Holes are only "minority carriers" in n-type germanium because the impurity atoms which donate most of the free electrons do not form useful "holes." When an electron breaks loose from one of the impurity atoms, the atom becomes ionized (positively charged), but it remains locked in place in the crystal structure and does not offer an available energy state into which electrons may move readily. The material remains electrically neutral when viewed as a whole, but the charge carriers which are free to move are predominantly negative, hence the n-type designation.

P-type germanium is formed by the addition of small amounts of an impurity such as indium which has only three valence electrons in the outer shell of each atom. When indium atoms are substituted for germanium atoms at some of the locations in a crystal lattice, each such atom becomes an "acceptor" for one of the bonding atoms from one of the neighboring germanium atoms, thus creating a vacancy or "hole" into which another electron may move. Each indium atom becomes negatively charged by the surplus electron attached to it, but these ionized atoms are unable to participate in current flow because they are locked in position in the crystal structure. Thus the holes (which behave like positive charges, although they are actually electron vacancies) are the majority carriers in p-type material, as shown by the energy diagram below.

The few electrons in the conduction band represent those that have acquired enough energy by thermal excitation to "escape" from the bonding role.

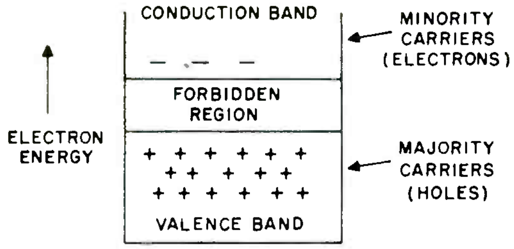

Figure 9

The conductivity of either n-type or p-type semiconductor material can be controlled by varying the concentration of the impurity material. For all practical purposes, each impurity atom donates one charge-carrying electron or hole, so the number of impurity atoms per unit volume is the primary factor in determining conductivity. Tunnel diodes use semiconductor materials with approximately 1,000-times more impurity atoms per unit volume than the materials used for conventional diodes and transistors. To complete this brief explanation of tunnel diode behavior, we should examine the properties of p-n junctions with both low and high concentrations of impurity atoms.

<u>Properties of Conventional Semi-Conductor Junctions</u>

When a junction is formed within a single crystal between p-type and n-type materials, a permanent electric field is established within the crystal, producing the energy level conditions shown in the diagram below.

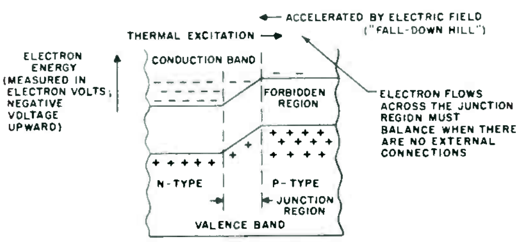

Figure 10

The surplus electrons in the conduction band of the n-type material can move across the junction rather freely to occupy some of the available energy states in the more sparsely-populated conduction band of the p-type material. At the same time, some of the holes in the valence band of the p-type material tend to pass through the junction to the less-populated valence band of the n-type material (this really means, of course, that electrons form the valence band of the n-type material move over to occupy some of the surplus vacancies at the same energy levels in the p-type material). The net flow of electrons from left to right continues only until the right side becomes negatively charged relative to the left by just the right degree to form an electric field across the junction great enough to maintain a right-to-left electron flow exactly equal to the left-to-right flow. The difference in charge raises the absolute energy level of the p-type material, and tends both to augment the p-to-n flow (by permitting the electrons traveling in this direction to "fall" to lower energy states) and to diminish the n-to-p flow (by requiring the electrons traveling in this direction to gain additional energy).

When an external voltage is applied to an n-p junction in the reverse direction (that is, with the n-type material positive), the height of the barrier in the energy diagram is soon increased to the point where virtually no electrons from the n-type region can acquire enough energy to cross the junction. Thus, there is no flow to cancel the flow of electrons from the p-side to the n-side, and a net current flows under the influence of the applied potential, as shown at the left in Figure 11. The magnitude of the current flow in the reverse-bias direction does not increase

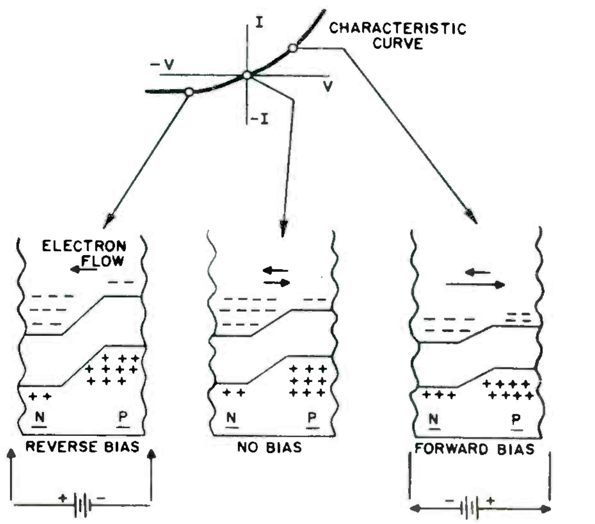

Figure 11

significantly with increasing voltage because of the very limited supply of minority carriers (thermally excited electrons and holes). When an external voltage applies a forward bias to the junction, as shown at the right in the sketch, the potential barrier is effectively reduced so that substantial numbers of the majority carriers can acquire enough energy to pass through the junction, permitting an appreciable current flow. When the forward bias becomes sufficient to "wipe out" the potential barrier, the junction becomes, for all practical purposes, a linear resistance--that is, the current changes become proportional to the voltage changes.

Tunnel Diode Junctions

The effective width of the junction region in a semiconductor diode is strongly influenced by the concentration of impurity atoms (i.e., the degree of "doping"), since the electric field in the junction region can be maintained only by the ionized impurity atoms locked into place in the crystal structure. If the impurity atoms are only sparsely distributed through the crystal, the space required to build up a sufficient net difference in charge to balance the electron flows is quite great, and the junction region is relatively wide. Conversely, a high concentration of impurity atoms, as in a tunnel diode, makes possible the creation of a very narrow or abrupt junction. At the same time, the high impurity concentrations in the two materials used in a tunnel diode lead to a considerable increase in the magnitude of the difference in potential required to balance the electron flow across the junction, as shown by the energy diagram below.

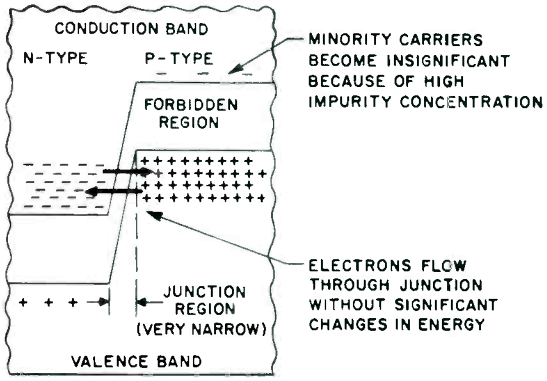

Figure 12

The combination of the very thin junction and the substantially greater offset in potential on the two sides of a tunnel diode junction makes possible the flow of electrons through the junction region by the process known as "tunneling." (The potential barrier is sufficiently great that relatively few electrons are able to cross the junction by acquiring enough energy to hurdle the potential barrier.) The great majority of the free electrons which impinge upon the junction region in the course of their random motions are reflected back into the same material from which they came, but a small fraction of them are able to pass through the junction and occupy some of the vacancies in the partially-occupied energy levels on the opposite side of the barrier. The probability that a given electron will pass through the junction is very low, but the supply of "charge carriers" is so great that even the small fraction of electrons which pass through can support an appreciable current flow. When no external voltage is applied, the electrons flow in both directions in sufficient quantities to balance each other. When a reverse bias is applied (n-type material made positive), there is a stronger net flow of electrons from the p-side to the n-side, corresponding to point (A) on the characteristic curve shown in Fig. 13.

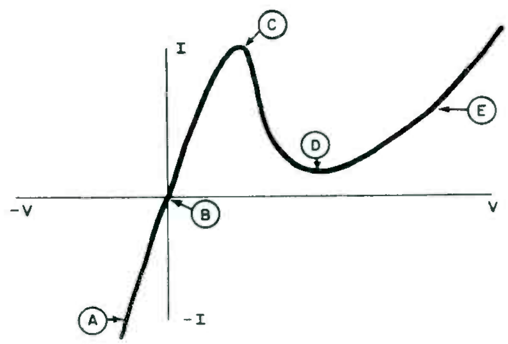

Figure 13

For slight amounts of forward bias in the region between (B) and (C) on the curve, the tunneling effect continues to operate and the junction conducts rather freely. As the forward bias is further increased in the range between (C) and (D), however, the current flow begins to decrease because the partially-occupied energy levels in the n-type and p-type materials are shifted to bring them opposite the forbidden regions on the opposite sides of the junction, thus removing the conditions which make possible the tunneling effect. At the current minimum, point D, the tunneling effect has virtually ceased, but a small current flow is possible because the potential barrier has been reduced sufficiently that some electrons are able to acquire enough energy by thermal excitation to cross the junction. As the forward bias is further increased to point E or beyond, the device behaves essentially the same as a conventional junction diode. As noted earlier, it is the negative resistance in the region from (C) to (D) that accounts for most of the tunnel diode's useful properties.

Materials for Tunnel Diodes

Some of the basic materials from which experimental tunnel diodes have been fabricated are listed below.

1. Germanium
2. Silicon
3. Gallium Arsenide
4. Indium Arsenside
5. Indium Antimonide

With any given material, the tunnel diode characteristics are controlled primarily by varying the concentration of impurity atoms or the area of the junction. The voltages at which the maxima and minima occur in the negative resistance characteristic tend to remain constant for a given material. For example, in all germanium tunnel diodes, the maxima and minima occur at about 50 and 350 millivolts, respectively, as shown in the following sketch.

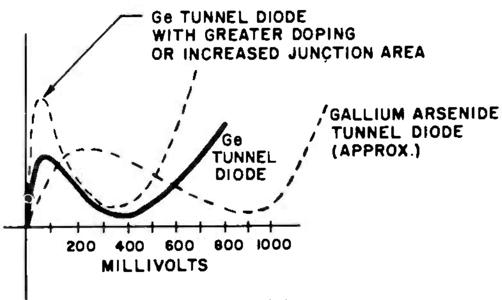

Figure 14

If the degree of doping on the junction area is increased, the current values change, but the curve retains essentially the same shape, as shown by the dotted line. Use of a different material, such as gallium arsenide, may move the voltage points, as indicated by the dashed line in the sketch.

Germanium will undoubtedly be a popular material for much of the early work with tunnel diodes, because its properties have been so thoroughly explored in conjunction with transistors and conventional diodes. Silicon does not appear particularly attractive for tunnel diodes because the effective "mass" of its charge carriers are somewhat greater than for the other materials, leading to impaired high frequency performance. One of the most promising materials is gallium arsenide, which not only offers a relatively wide voltage swing in the negative resistance region (approximately 1 volt), but also operates well at temperatures up to 400°C.

Typical Applications

Tunnel diodes show considerable promise for a wide variety of applications, including the following:

1. Oscillators (from low audio frequencies up to the kilomegacycle region).

2. Mixers or heterodyne converters.

3. Amplifiers.

4. A combined oscillator, mixer, and amplifier (all in a single diode!).

5. Bi-stable memory element for computers.

6. High-speed switching element for logic circuits.

7. Negative resistance element for filters.

8. Rectifier, voltage regulator, or small-signal detector. (Diodes for this service may be processed to minimize the first current peak, but will exploit the high conductivity in the reverse direction.)

Performance Characteristics

1. As amplifiers, tunnel diodes offer low noise and high gain-bandwidth products. In the microwave region, they are expected to compete closely with masers and parametric amplifiers with respect to low noise, and will be much lower in cost.

2. As computer elements, tunnel diodes offer extremely high switching speed, in the milli-microsecond region. The high switching speed and the broadband amplification capabilities result both from the extremely narrow junction and from the basic nature of the "tunneling" process, which is not greatly different from ordinary conduction in metals.

3. Tunnel diodes operate over a wide temperature range, extending from absolute zero up to possibly 400°C (for gallium arsenide units). This wide tolerance results from the fact that only majority carriers are used in the normal operating regions.

4. Tunnel diodes are relatively insensitive to atomic radiation (their tolerance is about 1,000 times greater than for most transistors).

5. In contrast to transistors, tunnel diodes will tolerate momentary overstress well beyond their normal operating conditions. In many cases, it is actually quite difficult to dissipate enough power in the junction to damage the device.

6. Both theory and preliminary experimental results indicate that tunnel diodes are relatively immune to contamination by moisture and atmospheric gases, thus eliminating much of the need for hermetic sealing.

7. Because hermetic sealing is possibly not required, tunnel diodes may be mounted in extremely small, light, encapsulated "packages," permitting even greater space and weight savings than transistors.

8. Tunnel diodes are inherently low-impedance devices with inherently high capacitance because of the "abrupt" junction. When biased for operation in the negative-resistance region, the equivalent circuit for a tunnel diode is approximately as shown below.

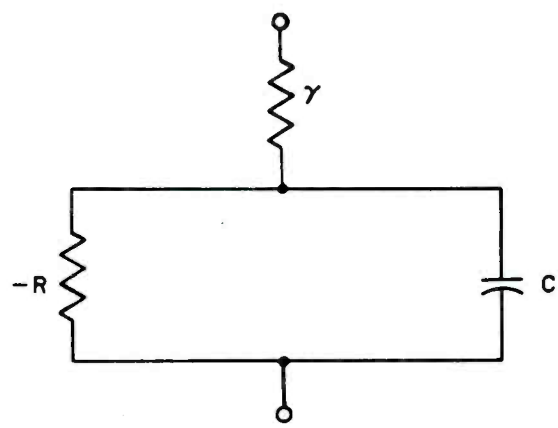

Figure 15

Typical values for the negative resistance (-R) might range from a fraction of an ohm to about 100 ohms, depending upon the material used, the degree of doping, and the junction area. The junction capacitance (c) likewise varies with the construction of the device, but is typically of the order of 75-100 micro-micro-farads. The resistor r represents the dissipative resistance of the ohmic contacts and the semiconductor material itself. r is usually small compared to -R, and can be neglected for many practical circuit analyses where first-order approximations are adequate.

9. The tunnel diode operates at a low power level, both with respect to input requirements and with respect to available output. In particular, the voltage swing is definitely limited in applications where operation must be confined to the negative-resistance region.

10. The two-terminal nature of the tunnel diode offers both advantages and disadvantages. On the plus side, most tunnel diode circuits are extremely simple, permitting significant reductions in size and weight and substantial improvements in reliability. On the other hand, the lack of inherent isolation between input and output is often a disadvantage making it impossible, for example, to cascade amplifier stages in the usual fashion. (It is expected, however, that new circuit approaches can be devised to overcome this handicap.)

11. Because of its inherently simple construction and relatively noncritical materials, there is good reason to believe that tunnel diodes can be produced at very low cost, once techniques for mass production have been developed.

12. The reliability of tunnel diode devices is expected to be substantially better than transistors because of the combination of factors listed above as items three through seven, plus nine and ten.

<u>Basic Circuit Concepts</u>

A comprehensive analysis of tunnel diode circuits is beyond the scope of this paper, but a few simple examples can be presented to illustrate the simplicity of tunnel diode circuits and to show how it is possible to perform a variety of functions with a two-terminal, negative-resistance device.

Low-Pass Amplifiers

A low pass amplifier may be constructed by connecting a tunnel diode, biased for operation in the negative resistance region, across the load resistance (R_L) of a signal source which may be symbolized by a constant current generator shunted by the source resistance (R_s). The bias resistor (R_b) shown in the following sketch is normally quite large compared to the load resistor, and can be ignored for first-order circuit analysis. Its function, of course, is to establish the DC operating point for the tunnel diode in the negative resistance region; it forms a voltage divider with the parallel combination of R_s and R_L.

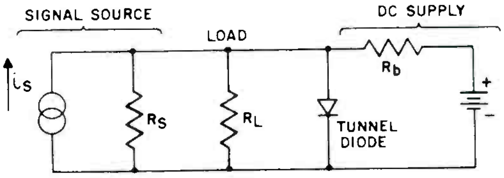

Figure 16

The voltage gain of this simple amplifier may be determined readily by comparing the performance of the circuit as shown with that of the same circuit with the tunnel diode removed. If the tunnel diode is not in place, the equivalent circuit becomes:

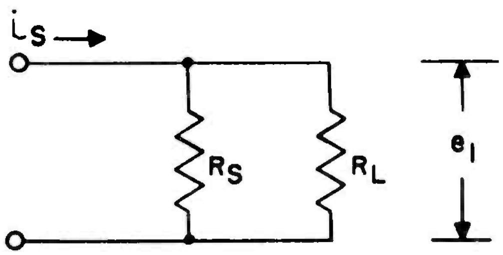

Figure 17

The output voltage under these conditions (e_1) is equal to the signal current multiplied by the equivalent resistance of R_s and R_L in parallel. In mathematical form,

$$e_1 = i_s \frac{R_s R_L}{R_s + R_L} = i_s R_T$$

where R_T is the equivalent resistance of the parallel combination.

When the tunnel diode is in place, the equivalent circuit becomes:

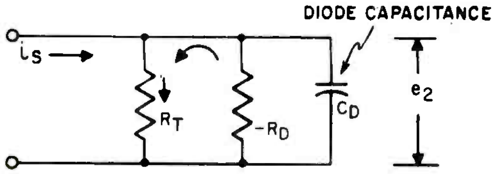

Figure 18

Because the diode acts as a negative reistance, it effectively becomes a current source, contributing additional current to the load and source resistors. The output voltage under these conditions (e_2) can be calculated as before:

$$e_2 = i_s \frac{R_T(-R_D)}{R_T - R_D}$$

The voltage gain of the amplifier is

$$\frac{e_2}{e_1} = \frac{-R_D}{R_T - R_D}$$

In theory, the amplifier offers infinite gain when the parallel combination of the source and load resistors equals the negative resistance of the diode, and the amplifier inverts the polarity of the signal if R_T is greater than R_D. In practice, it is desirable to keep R_T somewhat less in magnitude than R_D to avoid stability problems. (The simple biasing arrangement selected for illustration also requires that R_T be less than R_D.) When defined as the ratio of the power output with the diode in place to the power output with the diode removed and all other circuit constants the same, the power gain is simply the square of the voltage gain as expressed above.

The frequency response of the amplifier is limited by the diode capacitance (plus wiring capacitances) shunted across the effective output resistance. Although the capacitance may be of the order of 100 uuf, the resistance values are sufficiently low that the amplifier may have a gain-bandwidth product of 300 megacycles or better. (A voltage gain of 10 over a bandwidth of 30 megacycles is easily achieved!)

Band-Pass Amplifiers or Oscillators

The following circuit configuration can be used for either a band-pass amplifier or oscillator, depending upon the choice of circuit constants.

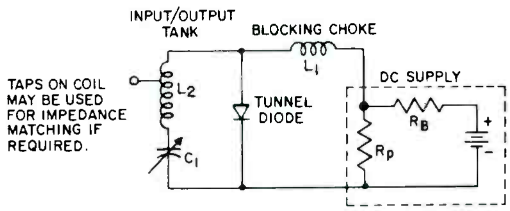

Figure 19

For analytical purposes, this circuit may be simplified as follows:

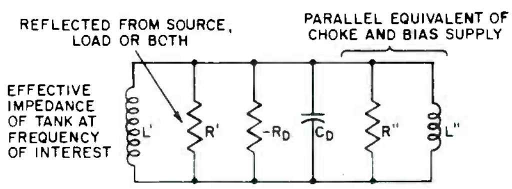

Figure 20

The circuit is equivalent to two tuned circuits tied together showing a common capacitor (the diode capacitance). The input/output tank circuit always appears inductive at the frequency of interest, because the diode capacitance is effectively in series with the tuning capacitor. To avoid oscillation in the DC supply circuit, the value of

$\frac{1}{R''} - \frac{1}{(R_D)}$ must be greater than $2\sqrt{\frac{C_D}{L''}}$.

If the circuit is to provide stable amplification, the value of

$\frac{1}{R'} - \frac{1}{R_D}$ must also be greater than

$2\sqrt{\frac{C_D}{L'}}$. If $\frac{1}{R'} - \frac{1}{R_D}$ is made less than

$2\sqrt{\frac{C_D}{L'}}$, circuit will perform as an oscillator.

Logic Circuits

The very simple circuit shown below is so versatile in performing logic functions that it is sometimes referred to as a "universal" logic circuit.

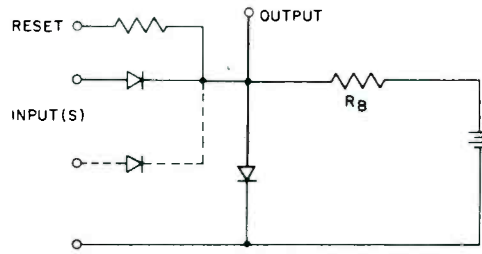

Figure 21

The bias resistor (R_B) is sufficiently large that the load line is almost horizontal, as shown below.

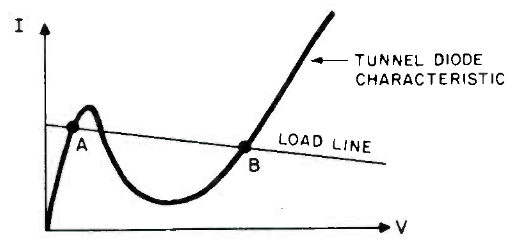

Figure 22

There are two stable operating points, A and B separated in voltage sufficiently that they may be used to serve as the two states of a binary memory. If the circuit is operating at point A, it can be switched to the B condition by the application of a positive voltage pulse of sufficient amplitude to push the voltage beyond the positive voltage peak; the circuit itself will then "flip" rapidly through the negative resistance region until stability is restored at operating point B. Likewise a switch from B to A may be initiated by a "reset" circuit which momentarily shorts the tunnel diode or applies a negative voltage of sufficient amplitude to shift the diode voltage below the current minimum into the negative resistance region. In some applications, the B to A switching could be accomplished by a <u>negative</u> pulse applied through a suitably-connected standard diode. In general, the circuit has the properties of a bistable flip-flop.

If an additional input connection is provided, the circuit can perform as a gate for either "and" or "or" service. For service as an "or" gate, the operating bias at point A should be adjusted so that a positive pulse on either of the two inputs will cause the circuit to switch to the B condition. For "and" service, the bias should be adjusted to lower the position of the load line so that it takes the presence of both input pulses simultaneously to cause the switching action to occur.

Physical Construction Techniques

Because of the very low impedances associated with tunnel diodes, it is necessary to employ construction techniques which minimize both the resistance and inductance of interconnecting leads. (Lead inductances can resonate with the diode capacitance to cause parasitic oscillations.) Tunnel diodes for experimental purposes have been constructed by soldering the semiconductor pellet between two flat plates measuring about 1/8 by 3/8 inch. Insulating spacers and encapsulants are used to hold the plates apart, forming a "sandwich" about 10 mils thick. This structure not only minimizes lead inductances, but also lends itself readily to circuits constructed with microstrip transmission line. The form factor for production versions of the device has not yet been standardized, but it is highly probably that some form of flatplate mounting will be used. In general, circuit construction practices will be similar to those currently used for UHF devices. Leads must be kept very short, and extensive uses should be made of flat strips, in preference to round wires, to minimize unwanted inductances. The small dimensions associated with tunnel diodes are highly compatible with the Micro-Module construction technique.

TRANSMISSION OF COLOUR TELEVISION SIGNALS

J. Davidse*
Philips Research Laboratories
N.V. Philips' Gloeilampenfabrieken
Eindhoven-Netherlands

Summary

The paper discusses the transmission of colour television signals according to the NTSC system. The choice of the chrominance signals, their bandwidths and of the subcarrier frequency is discussed. The consequences of the method of gamma correction and of deviations from the constant-luminance principle are considered. The significance of the statistics of the chrominance signal is pointed out.

1. Introduction.

The companion paper by de Vrijer ** explained the principles of colour television. It was stated that full information about the luminance and the colour of each part of the televised scene can be provided by three independent data. As a consequence the output of a signal source for colour television delivers three independent signals, which are commonly termed the red, green and blue signals and are denoted by the symbols R, G and B.

It is the task of the transmission system to combine these three primary signals in a suitable manner into one composite signal which can be transmitted by a radio frequency transmitter.

Before entering into the details of this transmission problem it seems worthwhile to survey the requirements to be met by the transmission system.
These are:

1. The colour-television receiver shall be able to present a good colour reproduction of the original scene.

2. For economical reasons the receiver has to be as simple as possible.

3. The transmission system has to be compatible with monochrome television that is: a normal black-and-white receiver tuned to the colour broadcast shall reproduce it as a normal black-and-white transmission. On the other hand a colour television receiver has also to be usable for monochrome transmissions.

4. In view of this compatibility requirement and in view of considerations of bandwidth economy the signal has to be such that it can be transmitted within the existing transmission channels. The planning of these channels is based on normal monochrome transmission; for the 625-line system their bandwidth is 7 Mc/s, the spacing between the vision and sound carriers being 5.5 Mc/s.

Hence about 5 Mc/s is available for the composite video signal.

At first glance these requirements must seem rather exacting and in some respects conflicting, but it will be shown that a good compromise is quite well possible.

2. Basic principles of colour-television transmission

2.1. Luminance and chrominance signals.

Let us first consider the nature of the information which has to be transmitted. In colour television we have the luminance and the colour of each part of the scene, while in monochrome television only information about the luminance is transmitted. As we have seen, full information concerning colour and luminance of the scene can be represented by three independent video signals. It will be clear that no information gets lost if we transmit three independent combinations of these primary colour signals instead of the signals themselves. More specifically it is feasible to choose one of these combinations in such a manner that it represents the luminance of the scene. The colour has then to be defined by two other combinations of the primary colour signals. Working this way and transmitting the luminance signal in quite the same manner as in monochrome television we gain two important advantages. First we fulfil our requirement of compatibility: a normal monochrome receiver will use the luminance signal in the normal way, hence a normal black-and-white picture is reproduced. Second we can use to good advantage a remarkable property of the human eye. A picture in which only the luminance

*Republished by Permission of Tijdschrift van het Nederlands Radiogenootschap where it was published in 1959 (pp. 255-272)
**Not included in this issue

information is displayed sharply whereas the colour information is displayed with much less sharpness is appraised as sharp by the eye. This property can easily be shown by a simple experiment; in fact it has already many other applications, e.g. the well-known coloured picture postcards of landscapes and beach-scenes. These are commonly normal black-and-white pictures to which very roughly some colour is added. The result may be liable to discussion from the aesthetic point of view, but its main drawback is certainly not a lack of sharpness.

If the three primary colours of the system are chosen as explained in the paper by de Vrijer, the luminance of the scene is given by the signal:

$$Y = 0.59G + 0.30R + 0.11B \qquad (1)$$

This signal is transmitted in the same way as a normal monochrome signal, i.e. it is transmitted with the full 5 Mc/s bandwidth. The two remaining signals determine only the colour to be reproduced and hence can be transmitted with a much smaller bandwidth according to the principle mentioned above.

2.2. The dot-interlace principle.

At this stage we must look for a method to find room for these two narrow-band signals in the video band which is seemingly already fully occupied by our luminance signal. For this purpose we can make use of the "dot-interlace principle". According to this principle the disturbing effect caused by a foreign signal in the video band is only small if the disturbing frequency is an odd multiple of half the line frequency, as is easily seen from Figure 1. In this figure representing part of a television scanning pattern with a disturbing frequency being present which is an odd multiple of half the line frequency, the letters A denote the dots produced by the disturbing signal in the first field of scanning. The letters B denote the dots occuring in the second field and in the same manner the letters C and D represent the dots occuring in the third and the fourth field. Hence a complete cycle of the disturbing pattern takes up two full frame scannings. As is easily seen from our figure the disturbing patterns are opposite each other in successive lines and in successive frames. Because of the integrating properties of the eye the light impressions of successive lines and frames due to the disturbing signal will more or less compensate each other. Hence a

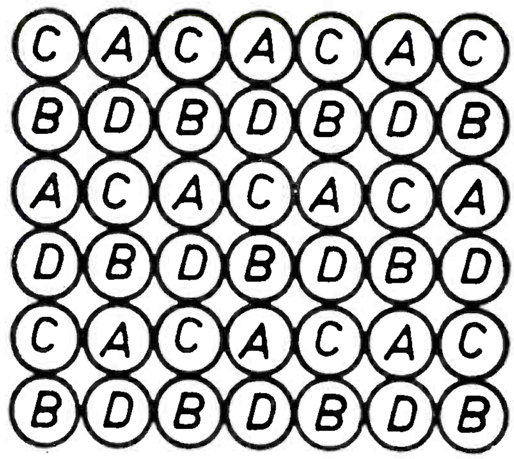

Fig. 1
Scanning pattern for odd multiple of half the line frequency.

spurious signal of such a frequency is only slightly objectionable. This enables us to introduce one or more subcarriers into the luminance signal, provided their frequencies are odd multiples of half the line frequency. Modulation of our colour signals onto such subcarriers allows us to transmit these signals within the frequency band of the luminance signal.

3. The NTSC-transmission system.

3.1. Modulation and demodulation of the subcarrier signal.

The method of introducing one or more subcarriers into the luminance signal in the manner described above is employed in almost all known experimental transmission systems for colour television. These systems differ in the number of subcarriers employed, the way they are modulated and the choice of the modulating colour signals. Rather than giving a survey of all transmission systems investigated until now we shall confine ourselves in this paper to a more detailed discussion of the most developed system which is in our opinion the best one. This is the NTSC-system[7]) which was developed in the U.S.A. in a combined effort of all leading industries in the field, who for this purpose created the National Television System Committee. Of course this system is adapted to the American black-and-white standard but the underlying principles can be applied in European versions of this system as well.

This system employs only one subcarrier modulated by both colour signals. We shall denote these signals by I and Q and for the present not consider in what manner they are composed of the primary colour signals.

Let the angular frequency of the subcarrier be ω. The subcarrier is amplitude-modulated by the first signal I, the subcarrier being suppressed. This can be achieved by employing a balanced modulator; at the output of the modulator we obtain the signal $I \cos \omega t$. The second colour signal Q is modulated in quadrature onto the subcarrier in the same manner, hence the signal $Q \sin \omega t$ is obtained. Adding these signals together we get the composite chrominance signal:

$$I \cos \omega t + Q \sin \omega t = \sqrt{I^2 - Q^2} \sin \left(\omega t - \arctan \frac{I}{Q}\right) \quad (2)$$

From this formula we see that we have modulated the subcarrier in phase as well as in amplitude by the colour signals. To demodulate the composite signal we multiply it in the receiver by $\sin \omega t$ and by $\cos \omega t$, respectively, and find:

$$(Q \sin \omega t + I \cos \omega t) \sin \omega t = \frac{1}{2} I \sin 2 \omega t + \frac{1}{2} Q - \frac{1}{2} Q \cos 2 \omega t \quad (3)$$

After filtering out the terms with double subcarrier frequency we get $\frac{1}{2}Q$.

In the same manner multiplication by $\cos \omega t$ yields:

$$(Q \sin \omega t + I \cos \omega t) \cos \omega t = \frac{1}{2} Q \sin 2 \omega t + \frac{1}{2} I + \frac{1}{2} I \cos 2 \omega t \quad (4)$$

Hence, after filtering we obtain $\frac{1}{2} I$.

To perform the above operations, which are commonly termed synchronous detecting, we have to produce in the receiver a subcarrier which is exactly synchronous with the subcarrier in the transmitter. For this purpose a synchronizing signal is introduced into the composite video signal, consisting of about 9 cycles of subcarrier frequency with known phase and amplitude. This synchronizing signal (usually named the "colour burst") provides the reference phase for the local oscillator in the receiver. It is applied in the line-blanking interval after the line-synchronizing impulse. Thus the composite video signal will assume the shape given in Fig. 2.

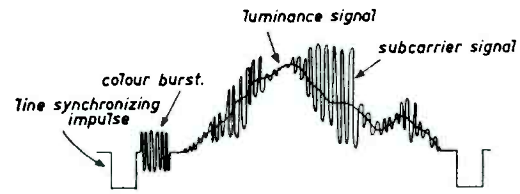

Fig. 2

Composite color television signal.

3.2. Choice of the colour signals.

3.2.1. General requirements. At this stage the problem of selecting the signals I and Q arises. First of all we have to bear in mind that in the receiver we have to retransform the signals Y, I and Q into the primary colour signals R, G and B. To get a simple transformation it is advantageous to choose for I and Q as well as for Y linear combinations of the primary colour signals. As we have seen Y is transmitted with full bandwidth whereas I and Q are narrow-band signals. Assuming equal bandwidths for the I and Q-signals to facilitate our considerations, we can split up each of the primary colour signals R, G and B into a low-frequency part, which is transmitted by the signals I and Q as well as by the luminance signal Y, and a high-frequency part which is transmitted by the luminance signal only. Denoting the low-frequency parts by the subscript L and the high-frequency parts by the subscript H, we can write:

$$Y = Y_L + Y_H = 0{,}30(R_L + R_H) + 0{,}59(G_L + G_H) + 0{,}11(B_L + B_H) \quad (5a)$$

$$I = \alpha_1 R_L + \beta_1 G_L + \gamma_1 B_L \quad (5b)$$

$$Q = \alpha_2 R_L + \beta_2 G_L + \gamma_2 B_L, \quad (5c)$$

where the symbols α and β denote numerical constants. In the receiver we obtain the low-frequency parts of R, G and B by weighted addition of Y_L, I and Q.

To the three primary colour signals obtained in this way we add the high-frequency part of the luminance signal Y_H, which is commonly termed the "mixed highs". Hence, to produce the primary signal R in the receiver we have to add to the luminance signal $Y_L + Y_H$ the signal: $R_L - Y_L$; to get B we have to add: $B_L - Y_L$ etc. However, it is desirable to avoid the necessity of having to form the separate parts Y_L and Y_H in the receiver, since this leads to rather severe complications. Formation of the signal Y_L from the luminance signal requires a filter whose amplitude response is exactly matched to that of the corresponding filters in the transmitter. This complication is avoided if the I and Q-signals are chosen in such a manner that it is possible to obtain in the receiver the signals $R_L - Y_L$, $B_L - Y_L$ and $G_L - Y_L$ without making use of the luminance signal Y, but using only the signals I and Q. Obviously to make this possible it is sufficient to choose for I and Q linear combinations of $R_L - Y_L$ and $B_L - Y_L$. If $R_L - Y_L$ and $B_L - Y_L$ are available we can obtain $G_L - Y_L$ from them quite simply, as this signal is already in itself a linear combination of $R_L - Y_L$ and $B_L - Y_L$.

Such a choice for the signals I and Q has an important second advantage. As was explained in the paper by de Vrijer the signal sources are arranged in such a manner that in reference white (Illuminant C) the three primary colour signals are equal in magnitude, hence for white we have $R = G = B = Y$. This means that for colourless parts of the scene $R - Y = B - Y = 0$ and hence also $I = Q = 0$. From expression (2) for the composite chrominance signal we see that the subcarrier amplitude then also equals zero. This means that we have no subcarrier signal in colourless parts of the picture, while the subcarrier amplitude increases with increasing saturation of the colour to be transmitted. The phase of the subcarrier signal depends on the hue of the colour. Assuming the transmission system to be linear and determining the contours of equal subcarrier phase in the chromaticity diagram we obtain Fig. 3.

The assumption of linearity of the system is an oversimplification as the actual system is on purpose made non-linear to correct for the non-linearity of the display tube; if we allow for this non-linearity in the calculation the final result differs considerably from the simple presentation of Fig. 3. In a subsequent part of this paper we shall discuss this problem in more detail.

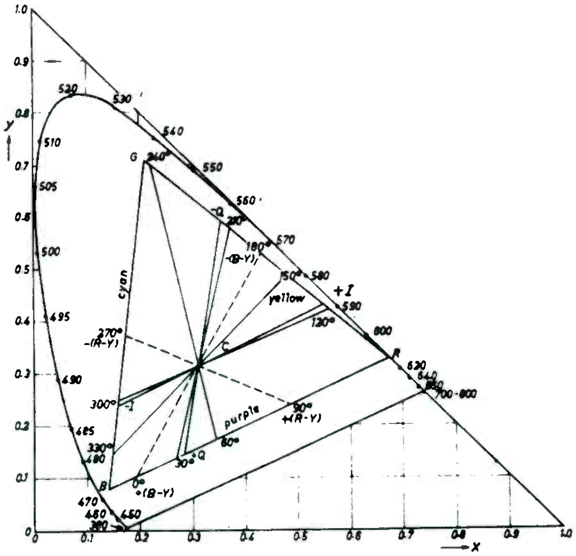

Fig. 3
Chromaticity diagram showing color-subcarrier phase as a function of reproduced chromaticity ($\gamma = 1.0$).

Returning to our discussion on the choice of the colour signals we shall have to deal with one problem remaining at this stage, i.e. the most suitable choice of linear combinations of R-Y and B-Y for the signals I and Q. When the signals I and Q are modulated with equal bandwidths onto the subcarrier, their composition is based on considerations concerning:

1. The allowable overswing of the subcarrier signal, that is: the amount by which the subcarrier signal surpasses the peak white and blanking levels of the signal.

2. The dependence of the hue of the colour to be reproduced on the phase of the subcarrier signal. The dependence has to be such that in all parts of the chromaticity diagram the proper relation exists between subcarrier phase and hue, viz. that the hue-differences which are just noticeable to the eye correspond to the same phase deviations.

For a detailed discussion of this problem we refer to the literature on the subject[10]. Investigations based on the considerations mentioned above lead to the following choice of both colour signals to be modulated in quadrature onto the subcarrier:

$0.49(B - Y)$ and $0.88(R - Y)$.

Hence, the composite colour-television signal can be written:

$$S = Y + 0.49(B - Y)\sin \omega t + 0.88(R - Y)\cos \omega t \qquad (6)$$

Fig. 4 shows the subcarrier amplitude and phase for saturated colours of maximum luminance if the colour signals are composed in this way.

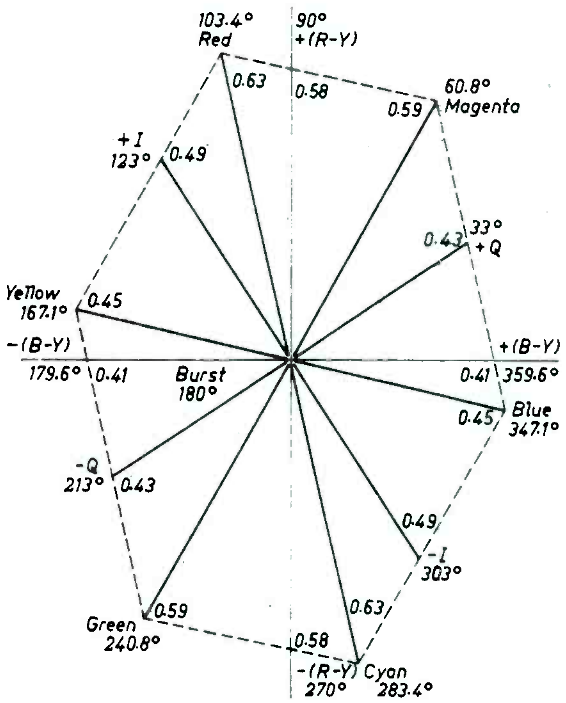

Fig. 4

Subcarrier amplitude and phase for saturated colors.

3.2.2. Unequal bandwidths of I and Q-signals. In the case of different bandwidths of the I and Q-signals the problem is more difficult. The same considerations as in the case of equal bandwidths apply, but in addition the question arises as to how the extra bandwidth has to be employed. Before dealing with this problem we shall first show that it is possible to transmit the signals I and Q with different bandwidths. This is easily seen if we observe more closely the process of synchronous detection, which is expressed in the relations (3) and (4). These relations indicate that for synchronous detection of the colour signals both sidebands of the modulated signals must be present. If one sideband of, e.g., the I-signal is suppressed we find full crosstalk of the I-signal into the Q-channel. Let us now consider the case of different bandwidths for the I and Q-signals. Let f_I be the cut-off frequency of the I-channel and f_Q that of the Q-channel. Furthermore we suppose $f_I > f_Q$. We now apply double-sideband modulation to the Q-signal but vestigial-sideband modulation to the I-signal in such a manner that all components of this latter signal which contain frequencies up to f_Q are double-sideband modulated, while the remaining components, containing frequencies between f_Q and f_I, are single-sideband modulated. In that case there will not be any crosstalk of the Q-signal into the I-channel, whereas only those components of the I-signal which contain frequencies between f_Q and f_I will crosstalk into the Q-channel. However, the Q-signal itself does not contain these frequencies because f_Q is the upper limit of its bandwidth, hence the crosstalking components from the I-signal can easily be removed from the Q-signal by a simple low-pass filter with cut-off frequency f_Q.

If the composite colour-television signal is composed along the lines set forth until now, its video-spectrum will be as presented in Fig. 5, where it is compared with that of normal monochrome signal, the only difference being the presence of the subcarrier signal.

Having thus proved the possibility of employing different bandwidths for the I and Q-signals we shall continue our discussion of the choice of these signals and their bandwidths.

It will be clear that in spite of the employment of dot-interlace a certain amount of crosstalk between the luminance signal and the subcarrier signal will be observable, due to the non-ideal integrating properties of the eye and the non-linear behaviour of the transmission system, which causes the subcarrier signal to be rectified. This remaining crosstalk is stronger as the subcarrier frequency is located lower in the video band and the bandwidths of the colour signals are larger. On the other hand it will be clear that a too severe bandwidth limitation of the subcarrier has to be avoided, too, as this will cause a lack of sharpness of the picture. Hence a good compromise has to be found between these conflicting requirements. This compromise depends among other things on the available video bandwidth, so it has to be determined anew for each individual adaptation of the NTSC-

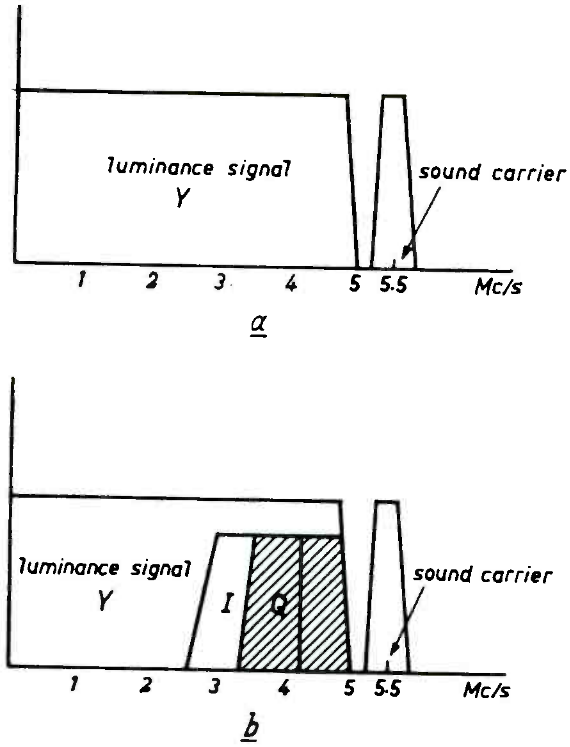

Fig. 5

(a) Monochrome-television channel.

(b) Color television channel.

system to an existing black-and-white standard. The only method to find optimum compromise is to carry out suitable experiments[2)3)7)11)]. Such experiments have to include an investigation into the effects of bandwidth limitation in the I and Q-channels on the quality of the reproduced picture, and into the mutual crosstalk between luminance and colour information.

The final result depends on, among other things, the non-linear behaviour of the transmission system. We shall therefore discuss the results of such experiments after the discussion of the non-linear behaviour of the system. For the present we shall give the final result of the experiments supposing the total available bandwidth to be 5 Mc/s, as is the case in the European 625-line system. A good compromise is obtained if the bandwidth of the I-signal is about 1.3 Mc/s and that of the Q-signal about 0.5 Mc/s, if we compose the I and Q-signals in the same manner as in the American system, that is:

$$I = - 0.28\ G + 0.60\ R - 0.32\ B \qquad (7a)$$

$$Q = - 0.52\ G + 0.21\ R - 0.31\ B \qquad (7b)$$

3.3. Block diagram of the complete transmission system.

We are now able to give a block diagram of the complete transmission system. The device which forms the composite NTSC signal from the three primary colour signals is commonly termed the encoder. Its block diagram is given in Fig. 6a. The input signals R, G and B are linearly transformed into the signals Y, I and Q by the matrix circuit. After having passed different low-pass filters the I and Q-signals are modulated in quadrature onto the subcarrier.

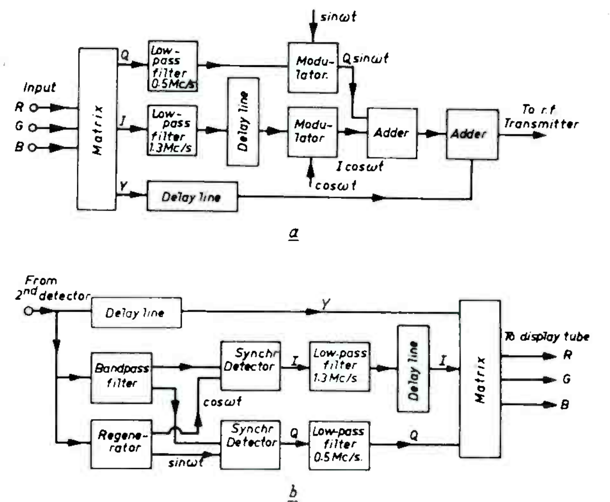

Fig. 6

(a) Block diagram of encoder.

(b) Block diagram of decoder.

The modulated signals are added together, thus forming the composite subcarrier signal, which is finally added to the luminance signal. In the luminance channel and in the I-channel delay lines are provided to match the different time delays of the various filters.

The composite signal is transmitted by a r.f. transmitter, received, and detected by the second detector in the receiver. The signal is then decoded, that is: from the composite signal the three primary colour signals are reformed. The decoder section of the receiver is given in Fig.6b. The subcarrier signal is taken from the composite signal with the aid of a bandpass filter. The burst signal is keyed out and employed in the regenerator which provides the demodulating subcarrier, which is fed to two synchronous detectors, producing the I and Q-signals from the composite subcarrier signal. After filtering these signals are fed to a matrix circuit together with the luminance signal Y. The output signals R, G and B of this circuit are fed to the display tube.

4. Gamma correction and non-linear behaviour of the system.

4.1 The constant-luminance principle.

For convenience we have so far ignored all non-linearities in the system. It will now be necessary to include this extra complication of the transmission problem into our considerations.

For a linear system the "constant-luminance principle" is valid. A colour-television system is said to be designed in accordance with this principle if none of the signals in the narrow-band chrominance channel contributes to the luminance of the reproduced picture. Among other things this means that the bandwidth limitations in the chrominance channel do not affect the luminance rendering[4)6)9)].

4.2. Gamma correction in colour-television.

As is well-known all display devices show non-linear characteristics. The relation between beam current and driving voltage is given by an expression of the type: $i = kV^{\gamma}$, where k denotes a numerical constant and γ is the "gamma" of the tube, which is usually about 2.2. If we want to avoid non-linear circuits in the receiver which correct this non-linearity, we have to introduce the correction at the camera end. In monochrome television this is accomplished in a simple manner by transmitting the signal $Y^{1/\gamma}$ instead of the signal Y, but in colour-television the matter is considerably more complicated. If we want to adhere to the constant-luminance principle, we must employ a luminance signal which carries all luminance information. The luminance of the original scene can be represented by:

$$L = 0.59\ G + 0.30\ R + 0.11\ B \tag{8}$$

Hence, to satisfy the constant-luminance principle the luminance signal has to be of the general type:

$$Y'' = p(0.59\ G + 0.30\ R + 0.11B)^{q}, \tag{9}$$

where p and q are constants. One might choose $q = 1/\gamma$, γ again being the gamma of the picture tube. In order to produce the correct picture this tube has to be fed with signals of the type: $G^{1/\gamma}$, $R^{1/\gamma}$ and $B^{1/\gamma}$. It will be clear that, supposing the luminance signal to be of the form (9), it is impossible to find I and Q-signals which, together with the signal Y", can be used to form the signals $G^{1/\gamma}$, $R^{1/\gamma}$ and $B^{1/\gamma}$, without requiring non-linear elements in the receiver. The application of non-linear elements has to be avoided, if possible, as it complicates the receiver severely. For this reason a compromise is accepted in practice: The gamma correction is applied to the primary colour signals, rather than to the luminance signal. This means that the signal sources are made to deliver directly the signals $G' = G^{1/\gamma}$, $R' = R^{1/\gamma}$ and $B' = B^{1/\gamma}$.
These signals are treated in quite the same manner as the signals R, G and B in the linear system as described above. Hence at the output terminal of the matrix network in the receiver we obtain the signals $G^{1/\gamma}$, $R^{1/\gamma}$ and $B^{1/\gamma}$,

which are precisely the signals needed to drive the picture tube. However, this implies that the luminance signal can be written:

$$Y' = 0.59G^{1/\gamma} + 0.30R^{1/\gamma} + 0.11B^{1/\gamma} \tag{10}$$

By comparison with (9) we see immediately that this signal does not satisfy the constant-luminance principle. Nevertheless the proper signals are fed to the picture tube, hence the proper luminance values are reproduced. This means that part of the luminance information has reached the display device via the narrow-band chrominance channel. Assuming $\gamma = 2$ to simplify the calculations, one finds that the reproduced luminance is given by:

$$L = Y^2 + 0.46I'^2 + 0.15I'Q' + 0.67Q'^2, \tag{11}$$

where Q' and I' denote the I and Q-signals according to (7) but composed of the gamma-corrected primary colour signals R', G' and B'.

As a consequence, the contours of equal subcarrier phase in the colour triangle will differ from those for the linear transmission system. They are presented in Fig. 7 for $\gamma = 2.2$, which appears on last page of this article.

4.3. Statistical properties of the chrominance signal.

From (11) we see that the subcarrier signal is elliptical, that is: the Q-signal contributes more to the luminance than the I-signal. Hence, we might expect from (11) that the bandwidth limitations in the Q-channel will have a greater bearing on the rendering of the luminance detail than those of the I-channel. Remarkable enough experiments show that in practice this holds only for artificial signals. If normal picture material is used the result is even opposite to what is to be expected from (11). A closer investigation into this phenomenon shows that it is caused by the statistical distribution of the colours in normal scenes. If we demodulate the subcarrier signal, according to Fig. 4, with different phases of the demodulating subcarrier thus obtaining different linear combinations of 0.49(B-Y) and 0.88(R-Y), then we find that the average signal content is maximum if the phase of the demodulating subcarrier is at an angle of about 30° with the (R-Y) phase, whereas a minimum occurs when this angle is about 100° larger. The average signal contents for the maximum and minimum axes are in the proportion of about 3:1. This result provides a sound reason for encoding the colour information in such a manner that the composing signals I' and Q' have different bandwidths. In that case the I-signal, which has the larger bandwidth, has to correspond to the colour information of the axis of maximum signal content as found above. The other signal has to be in quadrature with this signal and hence carries the colour information corresponding to an axis, close to the minimum signal-content axis.

4.4. The composite colour-television signal.

For reasons mentioned before the signal has to be such that the vector presentation of Fig. 4 applies, that is: R - Y and B - Y must be modulated in quadrature onto the subcarrier. However, this does not imply that the signals R - Y and B - Y themselves have to be used in the modulation process. As is readily seen from Fig. 4 an infinite number of sets of axes which are in quadrature to each other can be given. According to our experimental results the most suitable set of these axes is at an angle of about 30° with the set R' - Y'/B' - Y'. In the American system this angle is 33°. As was explained before the I-signal is vestigial-sideband modulated on the subcarrier while the Q-signal is double-sideband modulated. The composite signal can therefore be written:

$$S = Y' + Q' \sin(\omega t + 33^\circ) + I \cos(\omega t + 33^\circ) \qquad (12)$$

At frequencies below f_Q we have double-sideband modulation for I' as well as for Q', hence at these frequencies the given expression has to be equivalent with our original expression (6), or:

$$S = Y' + 0.49(B' - Y') \sin \omega t + 0.88\,(R' - Y') \cos \omega t. \qquad (13)$$

To be in accord with these equations the I' and Q'-signals have to be:

$$Q' = 0.41\,(B' - Y') + 0.48\,(R' - Y') = - 0.28G' + 0.60R' - 0.32B' \qquad (14a)$$

$$I' = -0.27(B' - Y') + 0.74(R' - Y') = -0.52G' + 0.21R' + 0.31B' \qquad (14b)$$

These equations are equivalent with equations (7), given earlier.

4.5. Bandwidth of colour signals and crosstalk phenomena.

The choice of different bandwidths for the I' and Q'-channels being motivated, we still have to find the optimum values for these bandwidths. This has to be done by carrying out suitable experiments. We shall not describe such experiments in detail[1][2][11] but confine ourselves to the most important results.

If we transmit the luminance signal and the subcarrier signal separately, thus eliminating all chances that crosstalk occurs between these signals, we can obtain data about the visible effects of bandwidth limitation in the colour channels. As already mentioned these effects include luminance errors as well as errors in the colour reproduction, due to the lack of constant luminance in the system. On the other hand experiments can be carried out which permit the investigation of the visible effects of mutual crosstalk between the luminance signal and the modulated subcarrier signal. As to the crosstalk of the subcarrier signal into the luminance channel: due to non-ideal integrating properties of the eye and the rectification of the subcarrier signal caused by the non-linear characteristics of the picture tube, we cannot avoid having to use a subcarrier suppression filter in the luminance channel of the receiver.

It turns out, however, that it is not necessary to suppress the subcarrier signal completely. If a relatively narrow band around the subcarrier frequency is removed by filtering, the remaining crosstalk caused by the sidebands of the subcarrier signal is only slightly visible, and less so as the subcarrier frequency is located higher in the video band.

As to the crosstalk of the luminance signal into the subcarrier channel: the high-frequency components of the luminance signal in the neighbourhood of the subcarrier frequency are treated by the subcarrier demodulators as sidebands of the subcarrier signal and hence give rise to spurious output signals of these demodulators. The magnitudes of these spurious signals may be assumed to increase roughly proportional with the bandwidth of the I and Q-channels. The experiments show that these spurious signals are mainly visible by their luminance contribution. We can see from (11) that the luminance contribution of signals in the Q-channel is considerably larger than that of signals in the I-channel. As the luminance contribution of the signals is approximately proportional to the square of their amplitude it is roughly proportional to the square of their bandwidth. On the other hand we have seen that the information content differs considerably for both colour signals. Combining the experimental results on the statistics of colour signals and on the phenomena of mutual crosstalk between luminance and chrominance signals we may draw the following important conclusion: <u>For either colour signal there exists an optimum bandwidth. If one increases the bandwidth beyond this value the useful information gained by increasing the bandwidth is of minor practical importance whereas the crosstalk of the luminance signal into the chrominance channel increases rapidly</u>.

Experimentally it was found that this optimum bandwidth is about 0.5 Mc/s for the Q-signal and 1.3 Mc/s for the I-signal. For the 625-line television standard a subcarrier frequency in the neighbourhood of 4.5 Mc/s is, therefore, to be preferred. The most suitable frequency in this region is 4.429687 Mc/s, which equals $567 \times \frac{1}{2}f_L$, f_L being the line frequency. Since $567 = (3^4 \times 7)$ does not contain larger factors we can use simple frequency dividers.

4.6. Visibility of the luminance errors.

Finally we shall discuss briefly the consequences of the lack of constant luminance in the system. As we saw before part of the luminance information is transmitted by the narrow-band chrominance channel. This causes distortions in the luminance rendering, viz. at sharp transients in saturated colours, generally known as "luminance notches"[9]. These luminance notches are very pronounced in artificial pictures, such as colour-bar signals. However, in normal pictures these effects are only seldom visible. This is again due to the statistical properties of the subcarrier signal.

From (11) one can easily see that the effects of constant luminance failure occur mainly in saturated colours, that is: for large values of the subcarrier signal for normal pictures demonstrate that only during 1% of the total time the subcarrier level surpasses half the maximum value. As the luminance errors depend roughly quadratically on the subcarrier amplitude, we can state that only during 1% of the time the errors exceed 25% of their maximum possible value.

From all these experimental results it will be clear that in the transmission system of colour-television signals the statistical properties of the signals play an important role.

<u>5. Tritanopia of the eye to small objects versus signal statistics</u>.

It should be noted that the discussion of the transmission system as presented in this paper differs in some respects from that given in most publications on the subject. The choice of the colour signals and of their bandwidths is usually described as related to the tritanopia (a kind of partial colour-blindness) of the human eye to small objects[3)8)9)]. However, the results of many experiments on the adaptation of this transmission system to the 625-line television standard have shown us that certain phenomena, which cannot be explained with the usual theory, fit in very well with the "statistical" approach. Of course we do not doubt the well-established fact of tritanopia of the eye to small objects; we merely believe that it does not provide an adequate explanation of the phenomena observed if bandwidth limitation is applied to the chrominance signals.

In many respects both theories lead to the same results, but in detailed discussions on the possibilities of improvements of the system both views lead to somewhat divergent conclusions.

References.

1. K. McIlwain: Requisite color band-with for simultaneous color-television systems. Proc. I.R.E. 40, 8 (1952), p. 909.

2. R. D. Kell and A. C. Schroeder: Optimum utilization of the radio frequency channel for color-television. R.C.A. review 14 (1953, p. 133.

3. G. H. Brown: The choice of axes and bandwidths for the chrominance signals in NTSC color-television. Proc. I.R.E. 42, 1 (1954), p. 58.

4. W. F. Bailey: The constant luminance principle in NTSC color-television. Proc. I.R.E. 42, 1 (1954), p. 60.

5. I. C. Abrahams: The "frequency interleaving" principle in the NTSC standards. Proc. I.R.E. 42, 1 (1954), p. 81

6. D. C. Livingston: Colorimetric analysis of the NTSC color television system. Proc. I.R.E. 42, 1 (1954), p. 138.

7. D. G. Fink: Color television standards. McGraw-Hill 1955.

8. K. McIlwain and C. E. Dean: Principles of color television. Wiley 1956.

9. D. Richman: Directions of improvements in NTSC color television systems. Proc. I.R.E. 44, 9 (1956), p. 1125.

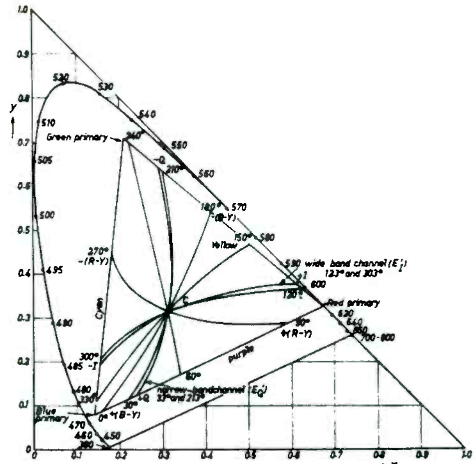

Fig. 7

Color-subcarrier phase as a function of reproduced chromaticity ($\gamma = 2.2$).

10. F. W. deVrijer: The choice of chrominance signals in the NTSC system with a view to the differential sensitivity of the human eye to colour. Acta Electronica 2 (1957-1958), p. 103.

11. J. Davidse: Versuche uber die Anpassung des NTSC-Farbfernsehsystems an die eropaische 625-Zeilen-Norm. NTZ 11, 9 (1958), p. 461.

STEREOPHONIC BROADCASTING
USING PULSE-AMPLITUDE MODULATION

H. F. Mayer and F. Bath

Rundfunktechn. Mitteilungen Jahrgang 3, 1959, pp. 174-179*

Summary

For transmission of the two stereo signals of a stereophonic radio broadcast, a time multiplex method operating with pulse amplitude modulation is proposed. The transmitting principle is described. The questions of compatibility, transmitter excitation, baseband width and radio band width, signal-to-noise ratio for single-channel and stereo reception, and synchronization are discussed. An account is given of the results obtained in trial workouts.

1 Introduction

With the rise of stereophonic records, stereophonic broadcasting gains in importance. As a matter of fact, the technical possibilities of stereophonic broadcasting are already the subject of intensive study in numerous countries. A considerable number of proposals for such broadcasting are under consideration.

The change of broadcasting to stereotransmission would be a step of greatest significance. Therefore it is necessary to select with greatest care the best method from among the many offered and to agree on this process before thinking of going on to the practical realization of stereobroadcasting. At least it should be required that the method be the same throughout Europe.

It is especially important, however, to take into consideration the existing areas of service of broadcasting stations, which must not be encroached upon by the introduction of stereophonic transmission.

The radio stations have therefore established the following demands for stereophonic broadcasting systems:

(a) They must permit transmission of the two stereophonic signals of the stereo program jointly over the same radio channel. That is, in the case of a UHF transmitter in a band of 300 kilocycles bandwidth.

(b) It must be possible for present receivers to receive stereo messages monophonically without any additional attachment. Vice versa, the stereoreceiver must be able to hear monophonic messages.

(c) The service areas which can receive monophonic messages in a frequency band must not be noticeably interfered with in their monophonic reception by the conversion of the radio station to stereobroadcasting.

(d) The decrease of the service area on passing from monophonic to stereophonic reception must be as small as possible.

(e) The radiation emission of the stereophonic program must not cause additional interference with other radio stations, or such interference must, at any rate, be very small.

(f) For economic reasons the additional cost for stereophonic receivers must be as small as possible. This holds for new receivers as well as for the conversion of existing ones.

(g) The service of stereo transmitters must be no less reliable than that of monophonic ones.

In the following a process is described which rests on well-known and tested foundations but has thus far met with little attention in stereotransmission, although it satisfies the foregoing methods very well--the process of using pulse amplitude modulation (PAM)

2 Transmission Procedure

The correct procedure for a stereo pickup is to place a microphone A to the right and a microphone B to the left of the studio's music platform. The microphone currents are amplified separately and carried to the transmitter by two separate circuits. They produce the two stereo signals A(t) and B(t) to be transmitted.

The principle of the transmission process is shown in Fig. 1. In the explanation of the mode of action we make use of the principles of the sampling theorem (Abtasttheorem), knowledge of which is assumed.

*Reprinted courtesy of Dr. Mayer and Dr. Bath, Siemens & Halske, A.G., Munchen (Germany) Translated from German by Mr. and Mrs. E. S. Allen of Iowa State University

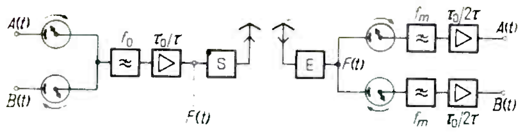

Fig. 1

Principles of stereobroadcasting

with pulse amplitude modulation.

At the outset the two stereo signals are sampled (abgetastet) by means of two electronic switches with period and in such a way that Signal B has the switch position of the phase opposite to that of A. The pickup frequency $f_o = 1/$ is twice the maximal signal frequency f_m occurring in A and B; that is, $f_o = 2f_m$ --e.g., $f_o = 30$ kilocycles. The two signals are merged behind the switches. The sum signal (Fig. 2) consists of temporally interleaved (ineinandergeschachtelt) pulse of the short duration and pulse amplitudes A_1 B_1 A_2 B_2 ... If these pulses are now sent through a low pass filter of cutoff frequency f_o, the pulses, as indicated in Fig. 3, flatten out into so-called (sin x)/x waves --however, in such a way that the positions of the individual pulses are unchanged. After an amplifier which compensates for the loss of signal caused by the scanning and filtering process one obtains a steady sum signal F(t), which has amplitudes 2A and 2B respectively at the original pulse positions. This sum signal which, interleaved in time, contains the full information of the two stereo signals A and B, is broadcast over the radio transmitter S.

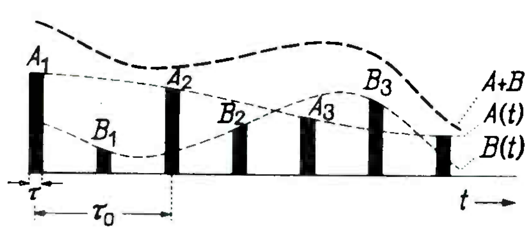

Fig. 2

Impulse sequence for pickup.

At the receiver the sum signal F(t) appears, e.g. at the output of the frequency discriminator of an UHF receiver. Here the path of transmission divides again into branches A and B. The two

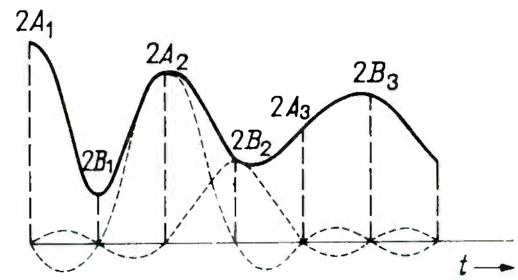

Fig. 3

Signal sum F(t).

electronic receiving switches run synchronously with the switches on the transmission side and pick up the sum signal F at the original pickup points. In this manner the result is achieved that the A impulses of the sum signal (Fig. 3) appear only in the upper branch, the B impulses only in the lower one. Low pass filters of the cutoff frequency f_m and amplifiers inserted later transform the short pickup impulses into the original stereo signals A(t) and B(t). Corresponding to the positions of the two microphones in the broadcast studio the A signal is conducted to the right loud speaker of a stereophonic receiver, the B signal to the left one.

By means of the temporal interleaving of the two stereo signals it is thus possible to transmit both signals over one transmitter. Here it must be particularly pointed out that the two transmission channels are fully equal in every respect. Every disturbance in transmission by noise, fading, etc. affects both signals in the same way. The reproduction through both channels thus remains symmetric under all circumstances. Unintentional interchange of A and B cannot take place.

3 Compatibility

For a monobroadcast the two signals A(t) and B(t) would already be united to a sum signal A+B in the studio,(Fig. 2) and would be radiated from the transmitter. In a stereobroadcast, however, the sum signal F(t) (Fig. 3) is radiated, which differs from the sum signal A+B as to signal frequency band and signal form. The question arises as to what an ordinary monophonic radio set receives if, in place of the monosignal A+B, the stereosignal F is sent to it.

Information on this is given most simply by the frequency spectrum of the stereosignal F(t). In Appendix 1 it is explained in greater detail that two processes are identical. In the first one, which is described, two signals A and B of 15 kilocycle band width each, by being picked up at the switching rate of 30 kc, are temporally interleaved and all frequencies above 30 kc are suppressed. In the second process the sum signal A+B is transmitted directly in the frequency range from 0 to 15 kc, and the difference signal A-B as lower side band of a suppressed carrier of 30 kc frequency.

Fig. 4 gives a schematic representation of the frequency spectrum of F(t). Each of the spectral amplitudes a and b represents an arbitrary partial vibration from the frequency spectrum of the signal A or B respectively. In the two sidebands the a components have the same direction, the b components opposite ones.

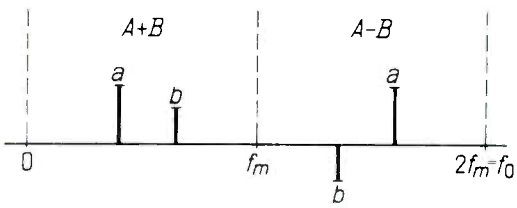

Fig. 4

Frequency spectrum of stereosignal.

A mono radio receiver which, in any case, transmits, in its low-frequency part, only the lower partial band up to 15 kc, thus reproduces the sum signal A+B, that is, the same signal which it would receive in a monobroadcast. The monoreceiver therefore perceives no distinction between mono and stereobroadcastings.

Of course one can also obtain monoprograms with a stereo receiver. In this case the monosignal A+B appears behind the discriminator E in place of the stereo signal F. The two switches distribute this signal equally to the A and B outlets. Obviously no stereo effect can occur here.

It is clear that both stereotransmission-monoreception and monotransmission-stereoreception are compatible in this process.

4 Frequency Deviation of the Transmitter

A further question arises as to whether, in a stereobroadcast, the frequency deviation of the FM transmitter must be made larger than in a monobroadcast. Since the carrier deviation is proportional to the momentary signal amplitude of the sum signal A+B or F respectively, this question amounts to asking whether the stereosignal F contains larger maximal amplitudes than the monosignal A+B.

From Fig. 4 it can be seen that the amplitude of signal F equals the amplitude of the sum signal A+B, plus that of the difference signal A-B. The stereosignal therefore has greater amplitude than the monosignal, and here, too, larger maximal amplitudes could be expected.

If, for instance, the two signals A and B were completely independent--that is, two time functions without any correlation--the amplitude of the stereosignal could be twice as great as that of the monosignal, since in such a case the amplitudes of the sum signal and the difference signal could be equal. Here, accordingly, amplitudes would have to be expected in the stereosignal larger, on the average, by the factor 2 than in the monosignal.

However, in a stereoprogram the two signals A and B are by no means independent of each other, since they originate from the same sound source. Large amplitudes, which are decisive for the frequency deviation, occur in practice only with low frequencies. And here one can calculate the modulation as if the maximal amplitudes were equal and were added in the sum signal and subtracted in the difference signal. Also for reasons of precaution against overcontrol the frequency deviation in monobroadcasting will be fixed at $f = 2A_m = 2B_m$. Larger amplitudes, however, do not occur even in the stereosignal F(t) (Fig. 3). It is reasonable, then, to assume that one can get along with the same frequency deviation for mono and stereobroadcasting.

5 The Transmitter Bandwidths

The bandwidth of the stereosignal F(t) is $2f_m = 30$ kc; the base band of each of the two stereo channels--and, of course, of the sum signal A+B--is 15 kc wide. If one wishes to undertake the interleaving of the two signals already in the studio, he needs a transmitter-studio link with a transmission band width of 30 kc, as against 15 kc formerly. If one undertakes the interleaving at the transmitter he needs two links of 15 kc each as against one link formerly. The transmission of stereo programs in the public news network requires either two normal broadcast lines (as compared to one hitherto) for each program or one new type of line which, in comparison with the former program lines, has the double band width.

The modulator of the transmitter must, correspondingly, be changed from 15 kc to 30 kc band width. Likewise the high frequency part of the transmitter and the antenna must be able to emit the broadened base band.

In view of the great dearth of frequencies in the medium wave range it will probably be impossible to double the width of the radio band of medium wave transmitters. In this case, then, the proposed procedure would not be applicable.

The conditions are more favorable in the UHF range. Usually the radio band width of an FM-UHF transmitter is calculated as B=2(f + f_m). Normal values in mono broadcasting are: Deviation f = 75 kc, maximal frequency f_m = 15 kc. This results in a band width of 180 kc. For a stereo broadcast, because of the double base band, the radio band width B equals 210 kc. It can be assumed that this slight broadening of the radio band will not cause any mutual disturbance of the UHF transmitters since the frequency separation of neighboring UHF transmitters normally amounts to 300 kc.

6 Signal/Sound Ratio at the Receiver

a) Monoreception of the stereoprogram.

As is well known in frequency modulation the spectral sound amplitude increases linearly with frequency and the spectral sound power per cycle of the frequency band increases quadratically. With the same frequency deviation, Fig. 5 is correct for both mono- and stereoreception, only with the difference that we are limited, in the case of monoreception, to the lower partial modulating band from 0 to f_m. Regardless of whether a monoprogram or a stereoprogram is broadcast, the signal and sound production in the lower part of the band are the same; that is, in monophonic reception of the stereosignal the signal/sound ratio remains the same as in the monophonic broadcast.

b) Stereoreception.

Here the conditions are considerably more complicated. The signal/sound ratio can be referred either to the sum signal F immediately behind the demodulator E or to the exit terminals A and B. In addition there is an acoustic signal/sound

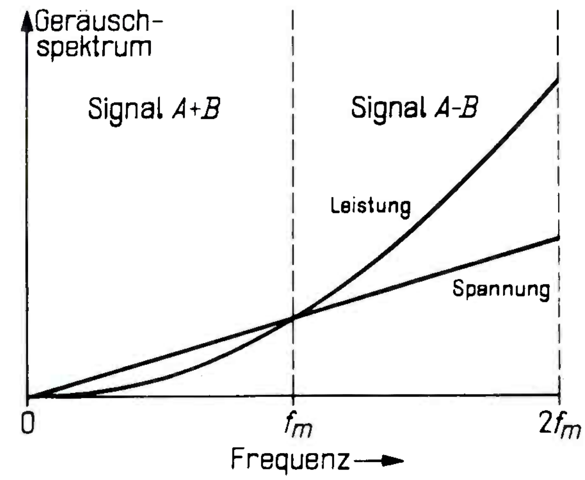

Fig. 5

Sound spectrum for frequency modulation

Sound spectrum (Gerausch-spektrum
Power (Leistung)
Amplitude (Spannung)
Frequency (Frequenz)

ratio, which the hearer experiences subjectively and which depends on the placement of the two loudspeakers, the position of the hearer relative to them, and on the complicated physiological processes involved in binaural listening.

We limit ourselves to the signal/sound ratio immediately behind the demodulator E (Fig. 1). Here the total signal amplitude (Fig. 4) and the total sound amplitude in the frequency band 0 to $2f_m$ (Fig. 5) must be taken into account.

The sound power, as integral of the spectral sound power over the frequency band is eight times as large as in the case of monophonic reception (frequency band 0 to f_m). If the two signals A and B were not correlated, the reception power with stereoreception would be twice that with monoreception. The signal/sound ratio would therefore deteriorate by a factor of 4.

With identical signals (A = B) A-B disappears. The upper partial band contributes only to the sound, not to the signal. Here, then, a deterioration by the factor 8 results. It is true that this is the most unfavorable case. Actually this factor will lie between 4 and 8.

7 Synchronization

On the transmitting side the two electronic switches are controlled by a frequency-stable pulse generator of recurrence frequency f_0 = 30 kc. Before the transmission there are superimposed on the two signals A and B direct voltages of different polarities, small in comparison with the signal voltages, so that the sum signal F(t) contains, besides the signal frequencies, also a carrier, in phase, of 30 kc and small amplitude. Fig. 6 shows the basic scheme of the circuit on the receiving side.

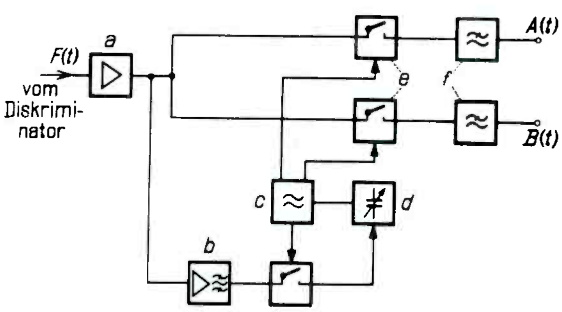

Fig. 6

Principle of synchronization for the receiver.

From the discriminator

a = Cathode follower
b = 30 kc filter
c = Pulse generator
d = Reactance tube
e = Electronic switch
f = Low pass filters

The carrier of 30 kc is filtered out by the filter b, amplified and used for the synchronization of a local pulse generator c which, in turn, controls the two electronic receiving switches as to frequency and phase. The whole circuit can be contained in a small box of dimensions 4 x 7.5 x 15 cm (Fig. 7).

The entrance to this box is connected with a discriminator and then furnishes at its two exits the two stereosignals A and B. The signal voltage delivered amounts to about 100 mv across 100 K . These are values such as are also furnished by the phonographic pickup of a player and are sufficient for the modulation of stereo cabinets.

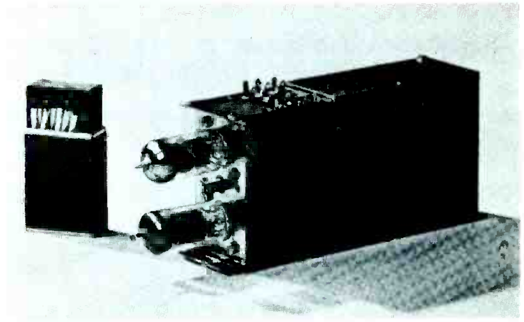

Fig. 7

Receiver.

Every radio receiver (UHF) can be equipped for the reception of stereo transmissions simply and economically by means of the supplementary box.

8 Practical Experiments

The procedure was tested first in the laboratory. A UHF test oscillator was used as a transmitter, different types of commercial radio sets as receivers. Commercial stereo cabinets of high quality were used for listening. The transmitter and the receiver were put in the same room for the tests. For the programs high quality stereotapes and stereorecords were used. It was possible to insert the transmission system (according to Fig. 1) between the two-channel program source and the play-back cabinet. By switches we were able to bypass the system, so that the program source was connected directly with the play-back cabinet.

The experimental setup was displayed to a group of experts of the federal postal system, of broadcasting, and of industry. In no case was it possible to distinguish between the transmission by the process described and the direct short-circuited transmission.

In these experiments it was, further, possible to switch mono and stereo-transmission on the transmitting side as well as on the receiving side, so that all possible cases of compatibility could be tested. It was ascertained that compatibility was assured in every respect.

On the basis of the good results the experiments are being continued.

Appendix 1.

Frequency spectrum of the stereosignal F(t).

The signal function in Fig. 2 can be thought of as produced by multiplication of the functions A(t) and B(t) with the impulse series $s_a(t)$ and $s_b(t)$ respectively (Fig. 8). In the case of a small ratio τ/τ_o the two impulse series can be represented approximately as follows, where $\omega_o = 2\pi/\tau_o$,

$$s_a(t) = \frac{\tau}{\tau_o}(1 + 2\cos\omega_o t + 2\cos 2\omega_o t + 2\cos 3\omega_o t + ..)$$

$$s_b(t) = \frac{\tau}{\tau_o}(1 - 2\cos\omega_o t + 2\cos 2\omega_o t - 2\cos 3\omega_o t + ..)$$

The mathematical representation of the impulse sequences according to Fig. 2 is therefore $A(t)s_a(t) + B(t)s_b(t)$.

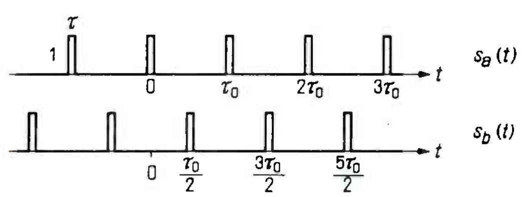

Fig. 8

Impulse trains.

If the stereosignals A and B are represented by two vibrations

$$A = a\cos(\omega_a t + \varphi_a) \quad \text{and}$$

$$B = b\cos(\omega_b t + \varphi_b)$$

then, if we use the abbreviations

$$\omega_a t + \varphi_a = \alpha, \quad \omega_b t + \varphi_b = \beta,$$

$$A(t)s_a(t) + B(t)s_b(t) = a\frac{\tau}{\tau_o}\left[\cos\alpha + \cos(\omega_o t - \alpha) + \cos(\omega_o t + \alpha) + ..\right] + b\frac{\tau}{\tau_o}\left[\cos\beta - \cos(\omega_o t - \beta) - \cos(\omega_o t + \beta) + ..\right]$$

The signal functions A and B appear in the original frequency position and in addition as lower and upper side bands of the carrier frequencies ω_o, $2\omega_o$, ...

Since, as can be seen from Fig. 1, the frequencies above $f_o = \omega_o/2\pi$ are suppressed, and since, furthermore, the factor τ/τ_o is concelled out by an amplifier, we obtain for the stereosignal F(t) (Fig. 3) the mathematical expression

$$F(t) = a\cos\alpha + b\cos\beta + a\cos(\omega_o t - \alpha) - b\cos(\omega_o t - \beta).$$

The frequency spectrum of F(t) (Fig. 4) consists therefore of a lower band between 0 and f_m, on which is superposed the signal

$$a\cos\alpha + b\cos\beta \ (=A + B),$$

and of an upper band between f_m and 2 f_m, on which is superposed the signal

$$a\cos(\omega_o t - \alpha) - b\cos(\omega_o t - \beta)$$

as lower side band of the carrier a_o).

For the pickup times t = 0, ,

2 , ... we obtain, indeed, F = 2A, and for the pickup times

t = /2, 3 /2, ...

the values F = 2B, as shown in Fig. 3.

Appendix 2.

Signal/sound ratio.

a) With reference to the stereosignal F(t).

We may assume that the two stereosignals A(t) and B(t) consist of a correlated part U and of uncorrelated parts X and Y respectively: i.e. A = U+X, B = U+Y. If, furthermore, the effective values which belong to these time functions are designated with the subscript o, then the power of the sum signal is

$$\overline{(A+B)^2} = \overline{(2U + X + Y)^2} = 4U_o^2 + X_o^2 + Y_o^2$$

The power of the difference signal is, correspondingly,

$$\overline{(A-B)^2} = X_o^2 + Y_o^2$$

If, further, V_o^2 is the sound power in the frequency band 0 to f_m, then the sound power in the frequency band f_m to $2f_m$ is equal to $7V_o^2$; the sound power in the total frequency band is thus $8V_o^2$.

Consequently the signal/sound ratio in the case of monoreception is

$$\left(\frac{S}{N}\right)_{mono} = \frac{4U_o^2 + X_o^2 + Y_o^2}{V_o^2} \tag{1}$$

The signal/sound ratio in stereoreception, referred to the exit at the demodulator E is

$$\left(\frac{S}{N}\right)_{stereo} = \frac{4U_o^2 + 2X_o^2 + 2Y_o^2}{8V_o^2} \tag{2}$$

In the case of complete correlation (X=Y=0) there results, then, a decrease by the factor 8 in stereoreception. In the case of completely uncorrelated signals (U=0) the resulting factor is 4. The truth lies, thus, between 4 and 8; on the average probably nearer to 8 than to 4.

b) With reference to the stereosignals A(t) and B(t)

It is possible to relate the signal/sound ratio to the two output signals A and B.

If one designates the sound voltage in the band from 0 to f_m with V_1 and in the band from f to 2f with V_2, whereby, in the preceding terminology,

$$\overline{V_1^2} = V_o^2 \quad \text{and} \quad \overline{V_2^2} = 7V_o^2$$

and if, further, one should, without picking up, obtain from the frequency spectrum of the stereosignal F(t) (according to Fig. 4) the two functions A+B and A-B by means of normal band separation and demodulation, he would obtain from the lower band the time functions (signal + sound)

$$A + B + V_1 = 2U + X + Y + V \tag{3}$$

and from the upper band

$$A - B + V_2 = X - Y + V_2 \tag{4}$$

By adding and subtracting it is possible to obtain A and B individually with the corresponding sound

$$A + \frac{V_1+V_2}{2} = U + X + \frac{V_1+V_2}{2}$$

$$B + \frac{V_1-V_2}{2} = U + Y + \frac{V_1-V_2}{2}$$

From this there result the following signal/sound ratios:

At the outlet of channel A

$$\left(\frac{S}{N}\right)_A = \frac{U_o^2 + X_o^2}{2V_o^2} \tag{5}$$

at the outlet of channel B

$$\left(\frac{S}{N}\right)_B = \frac{U_o^2 + Y_o^2}{2V_o^2} \tag{6}$$

Since we shall endeavor to transmit equal outputs in both channels, we have on the average $X_o^2 = Y_o^2$ in stereobroadcast. If one now compares the signal/sound ratio in the two stereo channels, that is, equations (5) and (6), with the ratio in the case of monoreception - equation (1) - then with complete correlation of A and B, a decreasing factor of 8 results , and with completely uncorrelated signals a decreasing factor of 4 to the disadvantage of stereoreception. Both ways of looking at it thus lead to the same result. These points of view, however, are probably too unfavorable, a fact which is immediately clear if one observes the output voltages at the two channels A and B - equations (3) and (4). The voltage component U in is phase in both signals and, when reproduced in the loudspeaker, will in a way be added in the listener. This acts , to this extent, in favor of a larger volume of sound than would correspond to the purely power-oriented addition of the individual signals in both channels. On the other hand, the larger share of sound voltage $V_2/2$ at the two loudspeaker terminals is in counter-phase - a fact which the ear will perceive as a weakening of the sound on the average.

It is therefore to be expected that the subjectively perceptible signal/sound ratio, combined from both stereo channels, will be better than the individual signal/sound ratios objectively measurable at the two loud speaker terminals.