VOLUME 26 SEPTEMBER, 1938 NUMBER 9

PROCEEDINGS

of

The Institute of Radio Engineers

Application Blank for Associate Membership on Page IX

Institute of Radio Engineers Forthcoming Meetings

ROCHESTER FALL MEETING
Rochester, N. Y.
November 14, 15, and 16, 1938

CLEVELAND SECTION
September 22, 1938

DETROIT SECTION
September 16, 1938

LOS ANGELES SECTION
September 20, 1938

NEW YORK MEETING
October 5, 1938

PHILADELPHIA SECTION
September 1, 1938

PITTSBURGH SECTION
September 20, 1938

WASHINGTON SECTION
September 12, 1938

PROCEEDINGS OF

The Institute of Radio Engineers

Volume 26 September, 1938 Number 9

CONTENTS

The Institute of Radio Engineers

GENERAL INFORMATION

INSTITUTE. The Institute of Radio Engineers was formed in 1912 through the amalgamation of the Society of Wireless Telegraph Engineers and the Wireless Institute. Its headquarters were established in New York City and the membership has grown from less than fifty members at the start to several thousand.

AIMS AND OBJECTS. The Institute functions solely to advance the theory and practice of radio and allied branches of engineering and of the related arts and sciences, their application to human needs, and the maintenance of a high professional standing among its members. Among the methods of accomplishing this is the publication of papers, discussions, and communications of interest to the membership.

PROCEEDINGS. The PROCEEDINGS is the official publication of the Institute and in it are published all of the papers, discussions, and communications received from the membership which are accepted for publication by the Board of Editors. Copies are sent without additional charge to all members of the Institute. The subscription price to nonmembers is $10.00 per year, with an additional charge for postage where such is necessary.

RESPONSIBILITY. It is understood that the statements and opinions given in the PROCEEDINGS are views of the individual members to whom they are credited, and are not binding on the membership of the Institute as a whole. Papers submitted to the Institute for publication shall be regarded as no longer confidential.

REPRINTING PROCEEDINGS MATERIAL. The right to reprint portions or abstracts of the papers, discussions, or editorial notes in the PROCEEDINGS is granted on the express condition that specific reference shall be made to the source of such material. Diagrams and photographs published in the PROCEEDINGS may not be reproduced without making specific arrangements with the Institute through the Secretary.

MANUSCRIPTS. All manuscripts should be addressed to the Institute of Radio Engineers, 330 West 42nd Street, New York City. They will be examined by the Papers Committee and the Board of Editors to determine their suitability for publication in the PROCEEDINGS. Authors are advised as promptly as possible of the action taken, usually within two or three months. Manuscripts and illustrations will be destroyed immediately after publication of the paper unless the author requests their return. Information on the mechanical form in which manuscripts should be prepared may be obtained by addressing the secretary.

MAILING. Entered as second-class matter at the post office at Menasha, Wisconsin. Acceptance for mailing at a special rate of postage is provided for in the act of February 28, 1925, embodied in Paragraph 4, Section 412, P. L. and R., and authorization was granted on October 26, 1927.

Published monthly by

THE INSTITUTE OF RADIO ENGINEERS, INC.

Publication office, 450–454 Ahnaip St., Menasha, Wis.

BUSINESS, EDITORIAL, AND ADVERTISING OFFICES

Harold P. Westman, *Secretary*

330 West 42nd Street, New York, N. Y.

INSTITUTE SECTIONS

ATLANTA—Chairman, C. F. Daugherty; Secretary, J. G. Preston, 230 Ansley St., Decatur, Ga.

BOSTON—Chairman, H. W. Lamson; Secretary, E. B. Dallin, 64 Oakland Ave., Arlington, Mass.

BUFFALO-NIAGARA—Chairman, G. C. Crom, Jr.; Secretary, E. C. Waud, 235 Huntington Ave., Buffalo, N. Y.

CHICAGO—Chairman, J. E. Brown; Secretary, E. Kohler, Jr., Ken-Rad Tube and Lamp Corporation, 2400 Sheffield Ave., Chicago, Ill.

CINCINNATI—Chairman, R. J. Rockwell; Secretary, M. M. Wells, Crosley Radio Corporation, 1329 Arlington St., Cincinnati, Ohio.

CLEVELAND—Chairman, L. N. Chatterton; Secretary, R. L. Kline, Winteradio, Inc., 3014 Lorain Ave., Cleveland, Ohio.

CONNECTICUT VALLEY—Chairman, D. E. Noble; Secretary, C. E. Keeler, 25 Webber St., Malden, Mass.

DETROIT—Chairman, E. H. I. Lee; Secretary, R. J. Schaefer, 9753 N. Martindale, Detroit, Mich.

EMPORIUM—Chairman, A. W. Keen; Secretary, M. C. Hoffman, Hygrade Sylvania Corporation, Emporium, Pa.

INDIANAPOLIS—Chairman, C. F. Wolcott; Secretary, B. V. K. French, P. R. Mallory and Co., Inc., E. Washington St., Indianapolis, Ind.

LOS ANGELES—Chairman, R. O. Brooke; Secretary, A. C. Packard, Station KNX, Columbia Broadcasting System, 5939 Sunset Blvd., Hollywood, Calif.

MONTREAL—Chairman, Sydney Sillitoe; Secretary, R. E. Hammond, 1216 St. Matthew St., Montreal, Que., Canada.

NEW ORLEANS—Chairman, G. H. Peirce; Secretary, D. W. Bowman, 8327 Sycamore St., New Orleans, La.

PHILADELPHIA—Chairman, A. F. Murray; Secretary, R. L. Snyder, 103 Franklin Rd., Glassboro, N. J.

PITTSBURGH—Chairman, W. P. Place; Secretary, R. E. Stark, 90 Pilgrim Rd., Rosslyn Farms, Carnegie, Pa.

ROCHESTER—Chairman, L. A. Du Bridge; Secretary, H. A. Brown, 89 East Ave., Rochester, N. Y.

SAN FRANCISCO—Chairman, Noel Eldred; Secretary, L. J. Black, 243–30th St., Oakland, Calif.

SEATTLE—Chairman, A. R. Taylor; Secretary, R. M. Walker, 2600-26th Ave., S.W., Seattle, Wash.

TORONTO—Chairman, R. C. Poulter; Secretary, N. Potter, Canadian National Carbon Co., Ltd., 805 Davenport Rd., Toronto, Ont., Canada.

WASHINGTON—Chairman, E. H. Rietzke; Secretary, L. C. Young, Naval Research Laboratory, Bellevue, Anacostia, D. C.

GEOGRAPHICAL LOCATION OF MEMBERS ELECTED AUGUST 3, 1938

Elected to the Associate Grade

Alabama	Huntsville, Radio Station WBHP	Davie, M. C., Jr.
Connecticut	Hartford, 157 Edgewood St.	McLeod, K. A.
District of Columbia	Washington, 1273 Brentwood Rd., N.E.	Black, R. R.
	Washington, 3115 Brothers Pl., S.E.	Kahler, F. C.
	Washington, 6-147 R St., N.E.	Stetson, R. B.
Illinois	Harvey, 14522 Clinton St.	Wiechmann, E. F.
Kentucky	Lexington, Radio Station WLAP	McCormack, J. E.
Maryland	Baltimore, 2115 Maryland Ave.	Roder, C. C.
Michigan	Detroit, 14810 Woodmont	Barrett, E.
New Jersey	North Arlington, 32 Beech St.	Green, N. H.
New York	East Islip, L.I., Second Ave.	Lang, J. G.
	Highland Mills	James, R. T.
Ohio	Dayton, Aircraft Radio Lab., Wright Field	Keto, J. E.
Oregon	Milwaukie, Route 11, Box 203	Tambara, H.
	Portland, 3928 S.E. Kelly St.	Bernard, J. D.
	Portland, Route 7, Box 1263	Brudwig, A. H.
	Portland, 1234 S.E. Tayler St.	Claypool, W. S.
	Portland, 3815 S.E. 28th Ave.	Clyne, W. E.
	Portland, 2272 S.E. 37th Ave.	Coon, W. B.
	Portland, Route 8, Box 156	Davis, C. B.
	Portland, 1955 N.W. Pettygrove	Fay, F. H.
	Portland, 4105 N.E. Shaver St.	Freitag, R. A.
	Portland, 4405 N.E. 22nd Ave.	Robinson, J. A.
	Portland, 6423 N.E. 9th Ave.	Spaulding, L. R.
	Portland, 1506 S.W. Montgomery St.	Steele, O. A.
	Portland, 633 N.E. Schuyler St.	Turner, H. E.
	Portland, 6036 N.E. 27th Ave.	Winkler, L. R.
Pennsylvania	Philadelphia, 4730 Pine St.	Batis, A. M.
Washington	Seattle, 2027 Eastlake Ave.	Gregory, C. M.
England	Birkdale, Lancs., 6 Blundell Dr.	Watts, R. V.
	Canterbury, Kent, 13 Station Rd. E.	Godden, A. W.
	Dulwich, London S.E. 21, 52 Park Rd.	Tanner, R. H.
	Kensington, London S.W. 10, 21 Tregunter Ave.	Wheeler, N. B.
	Welwyn, Garden City, 2 High Grove	Truscott, D. N.
India	Multan, c/o Dr. Wali Ram	Dara, S. N.
Italy	Milano, Via Plutarco 12	Danieli, G.
Scotland	Burghead, Morayshire, Phayrelands	Nicoll, W. L.
Switzerland	Zurich 6, Huttenstr. 58	Rusterholz, A. A.
Venezuela	Caracas, Ave. Norte 23[1] No. 25	Pachano, P. I.

Elected to the Junior Grade

Illinois	Chicago, 6148 S. Whipple St.	Jerolaman, W.
New York	New York, 3425 DeKalb Ave.	Aquila, M.

Elected to the Student Grade

California	Berkeley, 2538 Benvenue Ave.	Whaling, R. M.
	Walnut Grove, 2527 Hope St.	Hiehle, E. M.
Illinois	Chicago, 5305 Crystal St.	Sodaro, J. F.
Indiana	Angola, 403 W. South St.	Strickler, W. A., Jr.
	Angola, 320 S. Superior St.	Malheiros, L. G.
	West Lafayette, 116 Wiggins St.	Howe, R. M.
	West Lafayette, 414 Salisbury St.	Jackson, C. H.
	West Lafayette, 416 N. Main St.	Yadofsky, Z.
Massachusetts	Mattapan, 72 Clarkwood St.	Rosenthal, I.
Michigan	Detroit, 3883 St. Clair	Rhodehamel, A. J.
Oklahoma	Oklahoma City, 717 N. Lee	Woodward, O. M., Jr.
Oregon	Albany, 433 Denver St.	Dickson, F. H.
	Junction City, c/o E. L. Bryan, Route 3	Barclay, W. J.
Washington	Everett, 4201 Hoyt Ave.	Wells, W. W.

APPLICATIONS FOR MEMBERSHIP

Applications for transfer or election to the various grades of membership have been received from the persons listed below and have been approved by the Admissions Committee. Members objecting to transfer or election of any of these applicants should communicate with the Secretary on or before September 30, 1938. These applications will be considered by the Board of Directors at its meeting on October 5, 1938.

For Election to the Associate Grade

California	Bakersfield, Manly Apartments	Long, J. V.
	Sacramento, 515 Blackwood St.	Webster, N. D.
	San Rafael, 529 Mission Ave.	Walsh, C. E.
District of Columbia	Washington, 1008 Savannah St., S.E.	Congdon, C. L.
	Washington, U.S. Coast Guard, Treasury Dept.	Hopkins, A. F., Jr.
Florida	Miami, 3686 S.W. 25th Ter.	Roberts, F. G., Jr.
Illinois	LaGrange, 1505-41st St.	Dondanville, R. V.
Maryland	White Marsh	Francis, J. P.
Michigan	Bay City, 804 Wilson St.	Johnson, K.
New Hampshire	Laconia, Radio Station WLNH	Brooks, G. W.
New York	Valhalla, Box 101	Masterson, J. F.
Ohio	South Akron, R.D. 4, Box 507	Brennan, L. J.
Oregon	Portland, 8346 N. Dwight Ave.	Carruth, D. O.
	Portland, 2800 S.E. 74th Ave.	Cornell, W. R.
	Portland, Box 1807	Cutting, W. A.
	Portland, 7514 S.E. 17th Ave.	Denham, E. O.
	Portland, 2281 N.W. Everett St.	Dowers, A. A.
	Portland, R.D. 5, Box 811	Krieger, H.
	Portland, 3652 S.E. Stark St.	Loomis, D. D.
Pennsylvania	Glen Lyon, Box 633	Hoffman, W.
	Philadelphia, 3416 Baring St.	Gallagher, W. A.
	N.S. Pittsburgh, 801 Brighton Rd.	Troesch, J. M.
South Carolina	Clemson	Credle, A. B.
Wisconsin	West Allis, 2059 S. 75th St.	Osojnicki, J. J.
Australia	Sydney, N.S.W., c/o 2GB, Bligh St.	Cumpston, L. W.
Austria	Vienna, Fichtegasse 2	Mittelmann, E.
Canada	Toronto, Ont., 2520 Bloor St., W.	Patterson, M. C.
	Toronto, Ont., 584 Oakwood Ave.	Tonkin, H.
Canal Zone	Quarry Heights, Signal Office, Panama Canal Dept.	DeMarr, J. D.
England	Alfreton, Derbys., 23 Spa Croft, Tibshelf	Oxley, G. C.
	Birstall, Leicester, 20 Birstall Rd.	Griffin, N. B.
	Chelmsford, Essex, "Moulsey," Avenue Rd.	Crick, W. R.
	Chelmsford, Essex, 72 Longstomps Ave.	Edmunds, J. S.
	Chelmsford, Essex, 101 Broomfield Rd.	Masters, E. F. O.
	London N. 10, 30 Wilton Rd., Muswell Hill	Pamplin, S. F.
	Mitcham, Surrey, 53 Franklin Cres.	Payne, C. E.
South Africa	Johannesburg, c/o South African Broadcasting Corp., Box 4559	Olen, C. L.
South Australia	Ovingham, 27 Davenport Ter.	Allan, A. J.
South India	Cocanada, Suryaraopeta	Jogarao, C. V.
Tanganyika Territory	Daressalaam, Box 409	Sahota, R. S.

For Election to the Junior Grade

California	Los Angeles, 2919 S. Harvard Blvd.	Brenessel, A.
Switzerland	Geneva, 8 rue du Lac	Feyler, J.

For Election to the Student Grade

Michigan	Ann Arbor, 1032 Vaughn St.	Harunoglu, M. F.
	Flint, 2718 Chevrolet Ave.	Benson, A. E.
Germany	Muenchen, Arcisstrasse 48 Ir.	Zschirnt, H. H.
New Zealand	Auckland S.1, 16 View Rd., Mt. Eden	De Lisle, E. W.

OFFICERS AND BOARD OF DIRECTORS

(Terms expire January 1, 1939, except as otherwise noted)

President

HARADEN PRATT

Vice President

E. T. FISK

Treasurer

MELVILLE EASTHAM

Secretary

HAROLD P. WESTMAN

Editor

ALFRED N. GOLDSMIT

Directors

E. H. ARMSTRONG
J. E. BROWN
VIRGIL M. GRAHAM
ALAN HAZELTINE, *Past President*
R. A. HACKBUSCH
L. C. F. HORLE
E. K. JETT
A. F. MURRAY
B. J. THOMPSON

SERVING UNTIL JANUARY 1, 1940

H. H. BEVERAGE, *Past President*
RALPH BOWN
ALFRED N. GOLDSMITH
H. M. TURNER

SERVING UNTIL JANUARY 1, 1941

F. W. CUNNINGHAM
O. B. HANSON
C. M. JANSKY, JR.

Haraden Pratt
President of the Institute, 1938

Haraden Pratt was born in San Francisco, California on July 18, 1891. He received a Bachelor of Science degree in electrical and mechanical engineering in 1914 from the University of California.

He entered the amateur radio field in 1906. From 1910 to 1914 he did radio operating and installation work. From 1914 to 1915 he was assistant engineer for the Marconi Company of America at its high-power radio stations in California. From 1915 to 1920 he served as expert radio aide in the United States Navy Department serving on the Pacific coast and in Washington, D. C. From 1920 to 1922 he was in executive charge of the Federal Telegraph Company's radio construction and manufacturing activities at Palo Alto, California.

He had charge of the construction and operation of a high-frequency point-to-point telegraph system for Western Air Express in 1926 and 1927. During the next year he was in charge of the development of radio aids for air navigation at the Bureau of Standards, Washington, D. C. From 1928 to date he has been vice president and chief engineer of the Mackay Radio and Telegraph Company.

He has been closely identified with the standards work of the Institute since 1929, serving as chairman of the Standards Committee in 1935. He has been active on several radio committees of the American Standards Association.

He attended, as a technical advisor to the United States delegation or as a company representative, the International Radiotelegraph Conference held in Washington in 1927, the International Radio and Telegraph Conferences held in Cairo in 1938, and the conferences of the International Technical Consulting Committee on Radio Communication held in Copenhagen, 1931, and in Bucharest, 1927.

Mr. Pratt is registered as a professional engineer in the State of New York. He is a Fellow of the American Institute of Electrical Engineers, an Associate Fellow of the Institute of the Aeronautical Sciences, and a Fellow of the Radio Club of America. He became an Associate member of the Institute in 1914, a Member in 1917, and a Fellow in 1929.

Proceedings of the Institute of Radio Engineers
Volume 26, Number 9 *September, 1938*

TECHNICAL PAPERS

ASYMMETRIC-SIDE-BAND BROADCASTING*

By
P. P. Eckersley
(London, England)

Summary—*An economy in the frequency-channel width occupied by stations sending telephony signals can be made by removing one sideband from the transmitted spectrum. This operation however produces a distortion of the received signals which can only be minimized by an intensification of the carrier component compared with the side-band component. If this intensification is performed at the transmitter the carrier power must be enormously increased and the scheme would be impracticably uneconomic; it is furthermore impossible in broadcast technology to make the necessary alterations to intensify the carrier component in existing receivers all at once because they are publicly owned and extremely numerous. The distortion produced in the absence of carrier-wave intensification at either transmitter or receiver is mainly directly proportional to modulation. The modulation demanded in the transmission of ordinary speech and music is much less at the higher frequencies of modulation than that taking place in the lower middle frequencies. In order, therefore, to approach the ideal of the carrier- and single-side-band system, circuits have been devised to produce what is called an asymmetric side-band transmission in which only the outer parts of the side band are cut away. The distortion due to the cutting away of only the outer parts of one side band may be kept to negligibly small values because the intensity of the components in the outer parts of the side band, created by the higher frequencies of modulation, becomes less, since the modulating intensity is less at higher audio frequencies, as they are of greater frequency difference from the carrier.*

This paper gives an analysis relating the distortion produced by asymmetry between side-band components created by any given modulating component, the degree of distortion, and the depth of modulation. The expressions derived, and the assumption of a relationship between modulation frequency and modulation intensity, make it possible to derive "criterion" attenuation curves for a filter which, giving this performance, would not introduce more than a given constant and negligible distortion over a given range of modulation frequencies. These curves are based upon the assumption that phase asymmetry between counterpart side-band components is not so marked as to give a greater distortion than that inevitably introduced by magnitude asymmetry. It is therefore necessary to design a network, conforming to the requirements of differential attenuation of the side-band components, which does not introduce more than a negligible phase asymmetry between such components at values of side-band component frequencies where they have comparable magnitude. This network is of special form and is designed to combine the performances of band-pass and band-eliminator sections which are arranged to have, over the vital range of frequencies where the side-band components have comparable magnitude, inverse flex-

* Decimal classification: R550×R148. Original manuscript received by the Institute, April 2, 1938.

ures in their phase characteristics, and produce as a resultant a phase characteristic which does not introduce phase asymmetry between the differentially attenuated side-band components.

The side-band cutting filter produces a partial demodulation of the carrier and in order to preserve the level frequency characteristic given in symmetric transmission equalizer circuits are described designed to correct for the carrier demodulation.

A section is devoted to a study of the application of the system to practical uses and it is shown that if adopted several benefits would accrue which may be expressed as the alternatives (1) that if the present carrier-wave separations were maintained the interference known as "side-band splash" would be reduced tenfold, or (2) if the existing interferences were considered tolerable carrier-wave separations could be reduced to the order of 6 kilocycles, or (3) that all possibility of side-band splash would be eliminated, assuming receivers to be tuned asymmetrically, if separations were increased to 11 to 12 kilocycles.

A comparison is made between the system devised by the author and fully described in the paper and one, designed to fulfill the same objects, by N. Koomans.

It is suggested finally that the adoption of the asymmetric system might constitute a first step in an evolution the outcome of which would make it possible for transmitters to be adapted to the ideal system in which the carrier and only one side-band was transmitted.

List of Symbols

$a = M_N(1+\xi)/2$

$b = M_N(1-\xi)/2$

$d = 4\delta/M_N$

$d_0 = 4\delta_0/M_N$

f_m = mean pass frequency of band filter

f_s = frequency of side-band component

f_0 = carrier frequency.

f_1 = lower cutoff frequency of band filter.

f_2 = upper cutoff frequency of band filter.

Δf = frequency of modulating component.

Δf_∞ = frequency of modulating component at or above which α may be infinite.

M_A = modulation factor of carrier after passing through asymmetric filter.

$M_{A\xi}$ = modulation factor of carrier after passing through asymmetric filter when only magnitude asymmetry exists.

$M_{A\phi}$ = modulation factor of carrier after passing through asymmetric filter when only phase asymmetry exists.

M_N = modulation factor of carrier before passing through asymmetric filter.

M_ξ = magnification of amplifier associated with demodulation compensation network.

Δr = error in assuming $(x^2+y^2)^{1/2} = y + x^2/2$.

$r_1 = 1 - \Delta r$.

$r_2 = 1 + \Delta r$.
S = magnitude of one side-band-component vector.
t = time.
α = attenuation constant of filter designed to cut away one side band without the introduction of more than a constant and negligible distortion.
$\alpha_{e\xi}$ = attenuation constant of equalizer network for demodulation compensation when magnitude asymmetry is virtually zero.
δ = distortion factor, ratio of amplitude of second harmonic to fundamental.
δ_0 = constant value of δ chosen as negligible.
δ_ξ = value of δ when magnitude asymmetry is alone present.
δ_ϕ = value of δ when phase asymmetry is alone present.
θ = phase angle between currents entering and leaving filter.
θ_0 = value of θ at frequency f_0.
θ_a = value of θ at frequency $f_0 + \Delta f$.
θ_b = value of θ at frequency $f_0 - \Delta f$.
$\theta_1 = (\theta_0 - \theta_a)$ the result being positive.
$\theta_2 = (\theta_b - \theta_0)$ the result being positive.
ξ = ratio (<1) of amplitudes of two side-band components created by the same modulating component.
2ϕ = algebraic sum of θ_1 and θ_2, i.e., the algebraic sum of the phase angles made by the side-band components with the phase angle of the carrier component.
ψ = angle of vector resultant of two unequal-magnitude side-band vectors and the vertical.
$\Delta\omega = 2\pi\Delta f$.
$\omega_0 = 2\pi f_0$.
$\omega_s = 2\pi f_s$.

Introduction

THERE would be effected an economy in the number of channels required for the operation of a continental system of broadcasting if, according to systems common to commercial practice, only the carrier and one side band were transmitted.

The removal of one side band from the transmitted spectrum introduces, however, a distortion of the shape of the wave envelope of modulation which appears in the received product as a harmonic distortion of the fundamentals forming the modulations.

The harmonic distortion produced, when one side band is removed, is a function of modulation depth, the distortion increasing with increasing modulation. The effective depth of modulation of the carrier, and hence the distortion, can be reduced, while the received signals can be maintained at their original volume, if the carrier component is greatly increased above the value, in ratio to the modulation components, it normally possesses. If the carrier intensification takes place at the transmitter the power used must be greatly and wastefully increased beyond a value suitable for double-side-band transmission.

The economical solution to the problem of minimizing the distortion due to cutting away one side band, is to increase the value of the carrier component, without increasing the value of the side-band components, at the input to the receiver detector.

The introduction of such a system in broadcast practice would involve a radical alteration in the design of perhaps millions of receivers simultaneously and the solution of the problem to economize channel width in broadcast practice in such terms is seen to be, at any rate immediately, impracticable.

I. ASYMMETRIC-SIDE-BAND BROADCASTING

Asymmetric-side-band broadcasting is a name chosen by the author to describe a system in which the carrier and all of one side band is fully transmitted while the outer parts of the other side band are cut away.

This implies that for the lower frequencies of modulation a conventional and double-side-band system is employed and that the condition of carrier and single side band working is more and more approached as the frequency of modulation is higher.

The object of this system of transmission is to reduce harmonic distortion, due to the differential attenuation of side-band components, to negligible values. This object is achieved because the intensities of the upper-frequency components in the spectrum of sounds forming typical speech and music programs have a less intensity than the lower. By including both side-band components where modulation depth is large, and by cutting away more of the side band in regions where modulation depth is less, distortion may be kept constant and negligible.

This basic conception of asymmetric-side-band technique may be clarified by reference to Fig. 1 which shows the experimentally determined attenuation and phase characteristics of a constant-k band-filter section.

In asymmetric technique the carrier frequency f_0 is located nearer to f_1 the lower-frequency cutoff frequency of the filter than to f_m, its mean pass frequency, or nearer to f_2 than to f_m. If according to the conditions illustrated in Fig. 1, the modulation frequency be Δf_1 then the two side-band component frequencies are $f_0+\Delta f_1$, and $f_0-\Delta f_1$. But

Δf_1 is so low a modulation frequency that the attenuation of both side-band components is the same and virtually zero and both side-band components have the same magnitude and virtually the same phase relationship to the carrier. Thus if the modulation frequency is low there will be no side-band-component asymmetry and hence no distortion of the wave envelope of modulation.

If however, the frequency of modulation is increased to $\Delta f_2 > \Delta f_1$ then one side-band component of frequency $f_0 - \Delta f_2$ is considerably attenuated in relation to its counterpart component of frequency

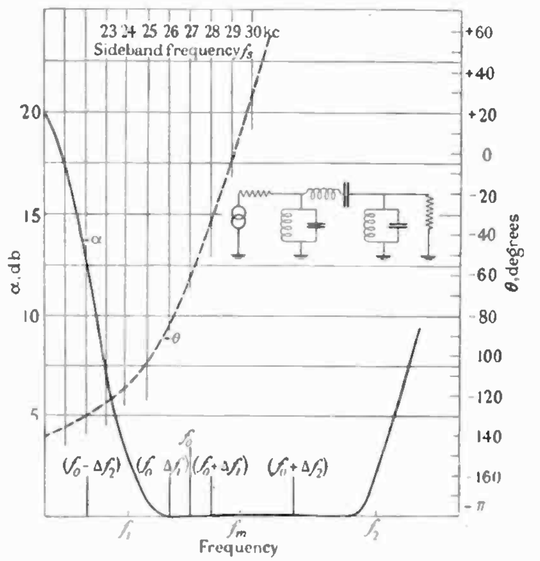


Fig. 1

$f_0 + \Delta f_2$. In this condition there will be asymmetry of magnitude between the two components as well as asymmetry of the phase differences between one side-band component and the carrier and the other side-band component and the carrier.

It will be proved hereafter that these asymmetries produce a distortion of the wave envelope, and hence a distortion in the received product, and that such distortions are a function of modulation.

But it is assumed, and can be proved experimentally, that in the transmission of typical speech and music programs the depth of modulation at a frequency $\Delta f_2 (> \Delta f_1)$ is less than that at a frequency Δf_1 and so the distortion introduced is smaller than if full modulation were demanded at the modulation frequency Δf_2.

Thus as the modulation frequency increases asymmetry increases but modulation depth decreases. The potentially increasing distortion due to increasing asymmetry is thus offset by the decreasing modulation and so the outer parts of the side band may be more and more cut away without the introduction of more than a negligible distortion.

The problem of designing a filter cutting away the outer part of one side band without introducing audible harmonic distortion can only be solved in terms of a knowledge of the relationships between the degree of asymmetry between any two side-band components created by a given modulation component, the magnitude of either of these components (which will be a function of modulation depth), and the distortion produced. Given this information it will then be necessary to fix upon some function to relate modulation depth with modulation frequency in the transmission of typical broadcast intelligence. Combining the information will allow us to derive expressions for the required attenuation-frequency and attenuation-phase characteristic of a filter which will cut away the outer part of one side band without the introduction of more than a constant and negligible distortion in the received product.

Distortion of the Wave Envelope Due to Side-Band-Component Asymmetry

Analysis of Modulation (Case of Symmetry)

Fig. 2 gives a vector diagram suitable for analyzing the process of the modulation of the intensity of high-frequency alternating currents at lower frequencies when both side-band components have symmetrical relationships to one another and to the carrier.

It is assumed throughout the following that the modulating component is sinusoidal and rectification "linear."

It is well known that the effect of the modulation of the intensity of currents having frequency f_0 by a component of lower frequency Δf is to produce two new components of frequencies $f_0+\Delta f$ and $f_0-\Delta f$.

Thus in Fig. 2 the vector OF is taken to represent a carrier, of unit magnitude, while the side-band component vectors are represented as $OA+OB$.

The vector OF is supposed to rotate at an angular velocity $2\pi f_0=\omega_0$ while OA rotates faster than ω_0 at an angular velocity $2\pi(f_0+\Delta f)=\omega_0+\Delta\omega$, and the other rotates slower than ω_0 by an angular velocity $2\pi(f_0-\Delta f)=\omega_0-\Delta\omega$. The projection on the vertical of the amplitude resultant of the three vectors OA, OB, and OF plotted against time gives, at that time, the amplitude of the modulated currents.

However, it is not necessary, when deriving the order of the distortion due to asymmetry, to have so complete an expression of the value of the modulated currents as would be given by plotting the high-frequency intensity at every instant; it is sufficient to find expressions for the shape of the envelope of the modulated currents. That is to say, if we can find the value of the maxima or minima of the currents at the instants of time when these occur a curve plotted through the values of such maxima or minima at the times when they occur will give the values of the audio-frequency components appearing after rectification.

Thus in Fig. 2(B) the carrier vector is reduced to rest and remains vertical while the side-band-component vectors are supposed to rotate in opposite directions at an angular velocity $\Delta\omega$. The resultant of the three vectors will then give the value of the maxima of the modulated currents ever occuring and the value of the maxima resultant plotted against time gives the shape of the wave envelope.

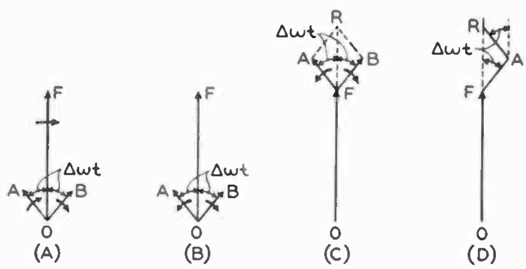


Fig. 2
——— = vectors
- - - - = construction lines.

It is furthermore clear that the wave-envelope shape, plotted about a horizontal line of height OF above zero, will give the value of the rectified currents.

In order to obtain an expression for the resultant of the three vectors the diagram Fig. 2(B) may be developed through Fig. 2(C) to Fig. 2(D).

Let the magnitude of each side-band component be S. The modulation factor, in a system where no distortion is present, may be defined as the ratio of the maximum increment of peak intensity when modulation is taking place to the unmodulated carrier peak intensity. If each side-band-component vector has a magnitude S then the increment of peak volts due to modulation is $2S$. Thus by definition and assuming the unmodulated carrier intensity to be unity the modulation factor

$$M_N = \frac{2S}{1} \qquad \text{or} \qquad S = \frac{M_N}{2}. \tag{1}$$

In Fig. 2(D) we require to know the resultant OR. If both side-band-component vectors are assumed to be vertical when time $t_0 = 0$ then at any time t they will each have moved in opposite directions of rotation through an angle $\Delta\omega t$. Clearly therefore OR is the sum of the

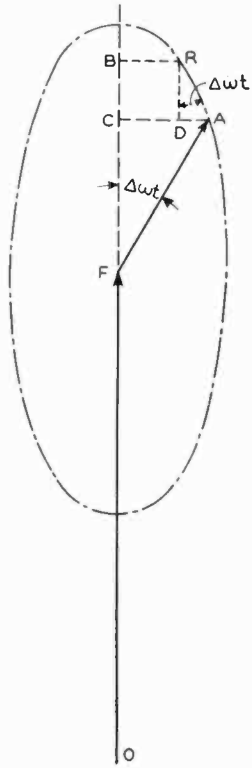


Fig. 3
——— = vectors
- - - - = construction lines
-·-·-· = locus of R.

unit-value-carrier vector and the projections on the vertical of $FA + AR$ or

$$OR = 1 + S \cos \Delta\omega t + S \cos \Delta\omega t$$

or from (1)

$$OR = 1 + M_N \cos \Delta\omega t. \tag{2}$$

The effect of rectification is to remove the carrier when it is seen that the low-frequency component appearing is a cosine curve having an intensity proportional to M_N, the modulation factor, showing that there is no distortion when both side-band components have symmetrical relationship to one another and to the carrier.

Distortion Introduced by Magnitude Asymmetry

Fig. 3 shows the effect of the introduction of magnitude asymmetry between the side-band components. One side-band component FA is supposed to remain at its original amplitude $S = M_N/2$ (due to a modulation intensity given by the modulation factor M_N) but the side-band-component vector AR is reduced in amplitude due to its differential attenuation and has a magnitude $\xi S = \xi\,(M_N/2)$.

ξ is thus the magnitude asymmetry factor and is always <1.

It is seen that

$$FC = FA \cos \Delta\omega t = \frac{M_N}{2} \cos \Delta\omega t$$

also

$$CB = DR = RA \cos \Delta\omega t = \frac{M_N}{2} \xi \cos \Delta\omega t$$

$$\therefore \quad FB = \frac{M_N}{2} (1 + \xi) \cos \Delta\omega t.$$

It is also seen that

$$BR = CA - DA$$

and since

$$CA = \frac{M_N}{2} \sin \Delta\omega t$$

$$DA = \frac{M_N}{2} \xi \sin \Delta\omega t$$

$$BR = \frac{M_N}{2} (1 - \xi) \sin \Delta\omega t.$$

Let

$$\frac{M_N}{2} (1 + \xi) = a \tag{3}$$

and

$$\frac{M_N}{2} (1 - \xi) = b. \tag{4}$$

Then,

$$FB = a \cos \Delta\omega t \tag{5}$$

$$BR = b \sin \Delta\omega t. \tag{6}$$

The locus of R is an ellipse and the magnitude of any semiaxis is given by

$$FR = (FB^2 + BR^2)^{1/2} = \left\{ a^2 \cos^2 \Delta\omega t + b^2 \sin^2 \Delta\omega t \right\}^{1/2} \tag{7}$$

and the angle of the axis to the vertical by

$$\Psi = \tan^{-1} \frac{b}{a} \tan \Delta\omega t. \tag{8}$$

The resultant OR is given by

$$OR = \left\{(1 + a \cos \Delta\omega t)^2 + b^2 \sin^2 \Delta\omega t\right\}^{1/2}. \tag{9}$$

The implications of these expressions will be dealt with hereafter when considering the case where both magnitude and phase asymmetry are present.

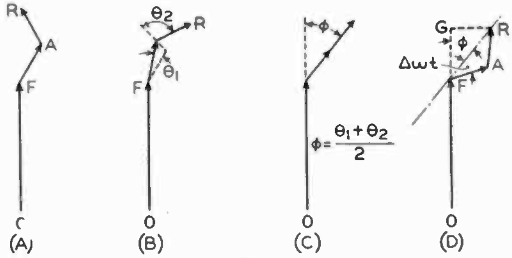


Fig. 4

——— = vectors
– – – – = construction lines
–·–·–· = locus of R.

Distortion Introduced by Phase Asymmetry

Phase asymmetry between side-band components created by any given modulation components may be introduced when, as in Fig. 1, the carrier frequency is not coincident with f_m. The phase characteristic of a filter is the phase angle between currents entering and leaving the filter plotted against frequency. Let θ_0 be this phase angle at the carrier frequency f_0, and θ_a and θ_b be the phase angles at side-band components corresponding to frequencies $f_0+\Delta f$ and $f_0-\Delta f$. Phase asymmetry is introduced when $\theta_0-\theta_a \neq \theta_b-\theta_0$ (θ_b being assumed $>\theta_0$, and θ_0 being assumed $>\theta_a$) that is to say when the algebraic sum of the phase angles made by the side-band components to the carrier is not zero or, in other words, is finite.

In Fig. 4(B) each side-band component, making in symmetric phase conditions an angle $\Delta\omega t$ to the vertical, is displaced by angles $\theta_1=\theta_0-\theta_a$ and $\theta_2=\theta_b-\theta_0$ where $\theta_1+\theta_2$ is finite. Rotating the vectors in opposite directions until they are in line Fig. 4(C) shows that since they are rotating at equal angular velocities in opposite directions the angle to the vertical they will make when in line is $\theta_1+\theta_2/2$ and that the locus of R is a straight line inclined at this angle $\theta_1+\theta_2/2$ to the vertical.

Let

$$\theta_1 + \theta_2/2 = \phi. \tag{10}$$

From the foregoing and considering Fig. 4(D) we can write that $FG = FR \cos \phi$ and from the analysis of the case for distortionless modulation $FR = M_N \cos \Delta\omega t$ so that $FG = M_N \cos \Delta\omega t \cos \phi$.

Also $GR = FR \sin \phi = M_N \cos \Delta\omega t \sin \phi$.

Evidently

$$OR = \left\{(1 + FG)^2 + GR^2\right\}^{1/2}$$
$$= \left\{(1 + M_N \cos \Delta\omega t \cos \phi)^2 + M_N{}^2 \cos^2 \Delta\omega t \sin^2 \phi^{1/2}\right\}$$

or

$$OR = \left\{M_N{}^2 \cos^2 \Delta\omega t + 2M_N \cos \Delta\omega t \cos \phi + 1\right\}^{1/2}. \tag{11}$$

The maximum and minimum values of OR are given by writing $\Delta\omega t = 0$ and 180 degrees, respectively, whence

$$OR_{\max} = \left\{M_N{}^2 + 2M_N \cos \phi + 1\right\}^{1/2} \tag{12}$$

$$OR_{\min} = \left\{M_N{}^2 - 2M_N \cos \phi + 1\right\}^{1/2}. \tag{13}$$

If $\phi = 90$, $OR_{\max} = OR_{\min}$. The implications of the result are explained hereafter.

Distortion Due to Magnitude and Phase Asymmetry in Combination

Fig. 5 illustrates the case when both phase and magnitude asymmetry are present. From the results of the foregoing it is clear that when $\xi < 1$ and ϕ is finite the locus of R is an ellipse having its major axis inclined at an angle ϕ to the vertical.

The values of FB and BR have been previously determined (see (5) and (6)).

From Fig. (5)

$$FG = FB \cos \phi = a \cos \Delta\omega t \cos \phi$$
$$GH = BK = BR \sin \phi = b \sin \Delta\omega t \sin \phi$$
$$FH = FG - GH = a \cos \Delta\omega t \cos \phi - b \sin \Delta\omega t \sin \phi.$$

Moreover

$$GB = HK = FB \sin \phi = a \cos \Delta\omega t \sin \phi$$
$$KR = BR \cos \phi = b \cos \Delta\omega t \cos \phi$$
$$HR = HK + KR = a \cos \Delta\omega t \sin \phi + b \cos \Delta\omega t \cos \phi.$$

The x and y co-ordinates of R are given by HR and $(1+FH)$ and we may write

$$x = a \cos \Delta\omega t \sin \phi + b \cos \Delta\omega t \cos \phi \tag{14}$$

$$y = 1 + a \cos \Delta\omega t \cos \phi - b \sin \Delta\omega t \sin \phi. \tag{15}$$

It is seen from the foregoing, where the three cases of (1) magnitude asymmetry, (2) phase asymmetry, and (3) magnitude and phase

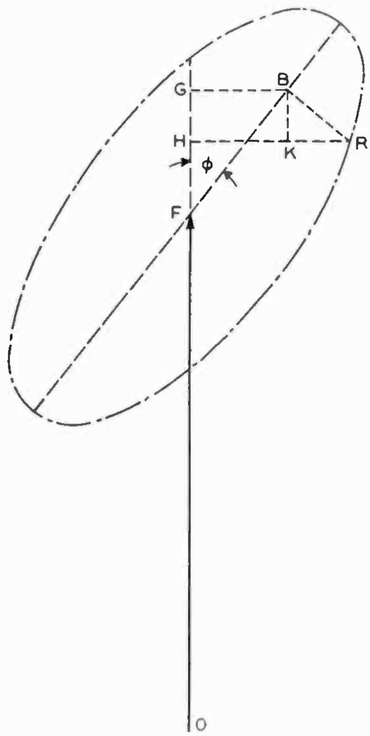


Fig. 5
——— = vector
– – – – = construction lines
–·–·– = locus of R.

asymmetry have been considered, that the resultant vector OR the magnitude of which plotted against time gives the shape of the wave envelope may, at certain instants of time, be different from the value it would have at those instants of time if neither magnitude nor phase asymmetry were present.

When magnitude asymmetry exists alone then, when $\Delta\omega t = 90$ or 270 degrees, the resultant OR is greater than it would be if symmetric modulation were present. The maximum divergence occurs when $\Delta\omega t = 90$ or 270 degrees. The occurrence of this divergence twice per cycle shows that the harmonic distortion is mainly of the form of the introduction of a new component having twice the frequency of the fundamental of the modulating component; i.e., the distortion introduces a second harmonic.

When phase asymmetry alone is present the maximum divergence of the value of OR from the value it would have in symmetric modulation is at $\Delta\omega t = 0$ or 180 degrees showing, once more, that the distortion is of the form of a second harmonic but displaced 90 degrees out of phase with the harmonic produced by magnitude distortion.

The combined effects of phase and magnitude asymmetry are seen also to produce mainly a second harmonic with phase intermediate between that introduced by magnitude or phase asymmetry depending upon which effect predominates.

When magnitude asymmetry was present then at a time $\Delta\omega t = 0$ or 180 degrees the magnitude of OR is $M_N(1+\xi)/2$ and $M_N(1-\xi)/2$ instead of, as in symmetric modulation, $(M_N/2)\times 2$ or $(M_N/2)\times(-2)$, i.e., M_N and $-M_N$, respectively. Thus when ξ is less than unity the effect of magnitude asymmetry is to produce, as well as a second-harmonic distortion, a demodulation of the carrier. The same effects are seen to occur both when phase asymmetry exists alone and when phase and magnitude asymmetry combine. Thus another form of distortion introduced by the asymmetric filter is to reduce the modulation factor of the carrier from M_N to M_A. The distortions are determined if, for the complete case, when $\xi < 1$ and ϕ is finite, an expression is derived giving the degree of the second harmonic and the degree of the demodulation.

In order to derive such expressions the value of OR must be expressed in terms of $\Delta\omega t$ to give the value of the fundamental and $2\Delta\omega t$ to find the harmonic. The value of the fundamental will then give M_A and the value of the terms in $2\Delta\omega t$ divided by the value of the fundamental, the distortion.

The value of OR is given by

$$OR = (x^2 + y^2)^{1/2} \tag{16}$$

and its complete expression would be given by substituting in (16) the values of x and y given in (14) and (15). It simplifies the algebra however to make an approximation.

This approximation depends upon two assumptions, as (1) that in the analysis of the vector diagrams representing the effects of asymmetry between side-band components the ratio x/y seldom exceeds a value to introduce any but a negligible error by writing

$$(x^2 + y^2)^{1/2} \cong y\left(1 + \frac{x^2}{2y^2}\right) = y + \frac{x^2}{2y}$$

and (2) that

$$y + \frac{x^2}{2y} \cong y + \frac{x^2}{2}\cdot$$

The latter approximation is justified because, in the cases analyzed, x is generally small compared with y when y is much greater or less than unity.

Thus the approximation chosen is that

$$(x^2 + y^2)^{1/2} = y + \frac{x^2}{2}\cdot \tag{17}$$

The degree to which the approximations may be relied upon is calculated hereafter.

From (14), (15), (16), and (17)

$$OR \cong 1 + a \cos \Delta\omega t \cos \phi - b \sin \Delta\omega t \sin \phi$$
$$+ \tfrac{1}{2}\{a \cos \Delta\omega t \sin \phi + b \cos \Delta\omega t \cos \phi\}^2.$$

Expanding the expression in the bracket gives

$$\tfrac{1}{2}\{a^2 \cos^2 \Delta\omega t \sin^2 \phi + 2ab \cos \Delta\omega t \sin \Delta\omega t \cos \phi \sin \phi$$
$$+ b^2 \sin^2 \Delta\omega t \cos^2 \phi\}$$

or

$$\tfrac{1}{4}\{a^2 \sin^2 \phi(1 + \cos 2\Delta\omega t) + 2ab \cos \phi \sin \phi \sin 2\Delta\omega t$$
$$+ b^2 \cos^2 \phi(1 - \cos 2\Delta\omega t)\}.$$

The value for OR can now be written in terms of the sum of three different forms; namely, (1) those not containing $\Delta\omega t$ and therefore a constant, (2) those containing $\Delta\omega t$ and therefore representing the fundamental, and (3) those containing $2\Delta\omega t$ and therefore representing the second harmonic.

Thus the approximation for OR may be written

$$[1 + \tfrac{1}{4}(a^2 \sin^2 \phi + b^2 \cos^2 \phi)] + [a \cos \Delta\omega t \cos \phi - b \sin \Delta\omega t \sin \phi]$$
$$+ [\tfrac{1}{4}\{a^2 \sin^2 \phi - b^2 \cos^2 \phi\} \cos 2\Delta\omega t + 2ab \sin \phi \cos \phi \sin 2\Delta\omega t].$$

The amplitude of the fundamental is thus given by

$$[a^2 \cos^2 \phi + b^2 \sin^2 \phi]^{1/2}. \tag{18}$$

The amplitude of the harmonic is given by

$$\tfrac{1}{4}\left\{(a^2 \sin^2 \phi - b^2 \cos^2 \phi)^2 + 4a^2b^2 \sin^2 \phi \cos^2 \phi\right\}^{1/2}$$

or

$$\tfrac{1}{4}(a^2 \sin^2 \phi + b^2 \cos^2 \phi) \tag{19}$$

whence the harmonic factor δ is

$$\delta = \frac{a^2 \sin^2 \phi + b^2 \cos^2 \phi}{4(a^2 \cos^2 \phi + b^2 \sin^2 \phi)^{1/2}} \cdot \tag{20}$$

Substituting the values for a and b in (20) a and b being given in (3) and (4) makes

$$\delta = \frac{M_N}{8} \frac{1 + \xi^2 - 2\xi \cos 2\phi}{(1 + \xi^2 + 2\xi \cos 2\phi)^{1/2}} \cdot \tag{21}$$

It is clear that the effect of the differential attenuation or phase shift of the side-band components is to reduce the effective modulation of the carrier at the output of the filter.

If M_A be the modulation factor at the filter output then clearly from (18)

$$M_A = \frac{M_N}{2}(a^2 \cos^2 \phi + b^2 \sin^2 \phi)^{1/2}$$

or

$$M_A = \frac{M_N}{2}(1 + \xi^2 + 2\xi \cos 2\phi)^{1/2}. \tag{22}$$

Special Cases

We may now derive expressions for the distortion for special cases. Thus suppose that $\phi = 0$ and $\xi < 1$, i.e., consider the case when magnitude asymmetry is alone existent, then from (21)

$$\delta_\xi = \frac{M_N}{8} \frac{(1 - \xi)^2}{1 + \xi} \cdot \tag{23}$$

If $M_N = 1$ and $\xi = 0$ (single-side-band transmission) the maximum value of δ_ξ is 0.125 or the harmonic distortion is 12.5 per cent, reducing in proportion to M_N. Also from (22)

$$M_{A\xi} = \frac{M_N}{2}(1 + \xi) \tag{24}$$

showing that if $\xi = 0$ (single-side-band transmission) and $M_N = 1$ $M_{A\xi} = 0.5$ showing that the effect of a filter, producing only magnitude asymmetry, is to demodulate the carrier and that for single-side-band working the demodulation is such as to halve the modulation at the input to the filter. This is obvious from first principles since in symmetric modulation the modulation factor is given by dividing the sum of the side-band amplitudes by the unmodulated-carrier amplitude and if one side-band component is removed the modulation is halved.

If $\xi = 1$ and ϕ is finite, i.e., when phase asymmetry is alone present, then

$$\delta_\phi = \frac{M_N}{8}\sqrt{2}\,\frac{(1 - \cos 2\phi)}{(1 + \cos 2\phi)^{1/2}}\,. \qquad (25)$$

Further in this special case

$$M_{A\phi} = M_N\sqrt{2}(1 + \cos 2\phi)^{1/2}. \qquad (26)$$

It is interesting to note that, although it is difficult to see how the condition could arise in practice, if the sum of the angles made by the side-band component vectors to the carrier were 180 degrees, i.e., when $2\phi = 180$ degrees, then

$$\delta_\phi = \infty \qquad \text{and} \qquad M_{A\phi} = 0.$$

This means that with equal magnitude vectors displaced by 180 degrees the carrier is completed demodulated and the distortion is infinite the resultant being purely of the form of a second harmonic.

The same result can be derived from (12) and (13) which show that when $\phi = 90$ degrees ($2\phi = 180$ degrees) $OR_{\max} = OR_{\min}$ and that the resultant is of purely second-harmonic form. Since (12) and (13) were not derived from an approximation and since (26) derived from the approximation gives the same result it is seen that expression (26) and hence expression (21) is accurate for values of $\phi = 90$ degrees.

Limits of the Approximation

It is now possible to show the limits of the approximation, i.e., at what values ξ, ϕ, and M_N the expressions (21) and (22) may be considered reliable.

The approximation given is

$$(y^2 + x^2)^{1/2} \cong y + \frac{x^2}{2}\,.$$

Let $100\,\Delta r$ be the percentage error in using the approximation then either

$$\Delta r = \frac{(x^2 + y^2)^{1/2} - y + \dfrac{x^2}{2}}{(x^2 + y^2)^{1/2}}$$

or

$$\Delta r = \frac{\left(y + \frac{x^2}{2}\right) - (x^2 + y^2)^{1/2}}{(x^2 + y^2)^{1/2}}.$$

These expressions can be rewritten as

$$(x^2 + y^2)^{1/2}(1 - \Delta r) = y + \frac{x^2}{2}$$

and

$$(x^2 + y^2)^{1/2}(1 + \Delta r) = y + \frac{x^2}{2},$$

respectively.

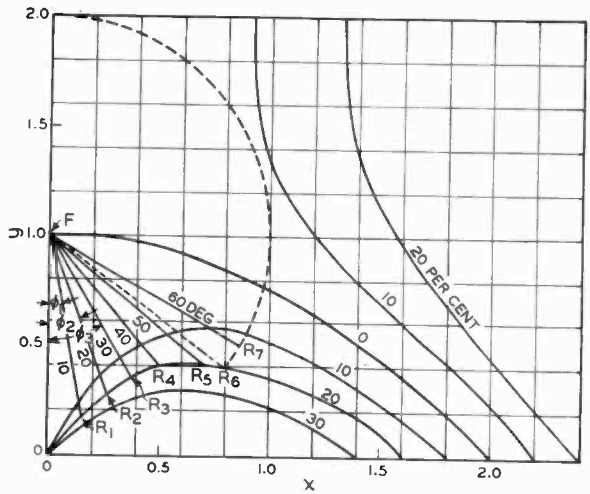


Fig. 6

If $(1-\Delta r)=r_1$ and
$(1+\Delta r)=r_2$ then
either

$$y = \frac{x r_1 \left\{\frac{x^2}{4} + 1 - r_1^2\right\}^{1/2} - \frac{x^2}{2}}{1 - r_1^2}$$

or

$$y = \frac{\frac{x^2}{2} \pm \left\{\frac{x^2}{4} + 1 - r_2^2\right\}^{1/2}}{r_2^2 - 1}.$$

The curves of Fig. 6 are plotted for different constant values of $100\,\Delta r$, the percentage error, against x and y, between values 0 and 2, respectively.

The point F in Fig. 6 can be taken to represent the tip of the carrier-amplitude vector, and lines FR_1, FR_2, etc., drawn at angles ϕ_1, ϕ_2 etc., as the major axes of the ellipses forming the locus of R. Let the lengths FR_1, FR_2, etc., $= l_1$, l_2, etc. Evidently

$$l_n = \frac{M_N}{2}(1 + \xi) \tag{27}$$

whence

$$\xi = \frac{2l_n}{M_N} - 1 \tag{28}$$

where l_n is any value of l.

But in Fig. 6 the lengths l_1, l_2, etc., are drawn between F and a line of constant error Δr_0. Thus (28) will give, for chosen values of FR and hence ϕ the value of ξ and M_N at which the error introduced will never exceed Δr_0. Δr_0 in Fig. 6 is chosen as 20 per cent.

Curves of Constant Distortion Plotted Against ξ and ϕ

It is now proposed to derive from (21) values of cos 2ϕ in terms of δ, ξ, and M_N in order that curves may be drawn for constant values of δ against 2ϕ as ordinates and ξ as abscissas.

This conversion, which involves straightforward, if somewhat heavy, algebra shows that we may write

$$\cos 2\phi = \frac{1 + \xi^2 + 2d\left[d \pm \left\{2(1 + \xi^2) + d^2\right\}^{1/2}\right]}{2\xi} \tag{29}$$

where

$$d = \frac{4\delta}{M_N}. \tag{30}$$

It is useful in plotting the required curves to find the values of ξ when $\phi = 0$ degrees and $\phi = 180$ degrees.

Thus when $\phi = 0$ we may use (23) and convert that expression to give

$$\xi_{\phi=0} = (1 + d) - \left\{d(d + 4)\right\}^{1/2}. \tag{31}$$

When $2\phi = 180$ degrees we may use (21) and write

$$\xi_{2\phi=180^\circ} = \left\{d(d + 4)\right\}^{1/2} - (1 + d). \tag{32}$$

Equation (25) may also be converted to give cos 2ϕ when $\xi = 1$ whence

$$\cos 2\phi_{\xi=1} = 1 + d^2 - d(d^2 + 4)^{1/2}. \tag{33}$$

Thus Fig. 7 shows curves plotted from (29), (31), (32), and (33) which gives curves of constant-percentage distortion plotted against ξ and ϕ for $M_N = 1$.

But it has been shown that the formulas are based upon an approximation. The limits of the approximation have been determined (see Fig. 6). The curves shown in dot-dash lines can be derived from (28) and drawn as in Fig. 7, as shown, to delimit the values of ξ and ϕ above which the error in using the curves of constant distortion is

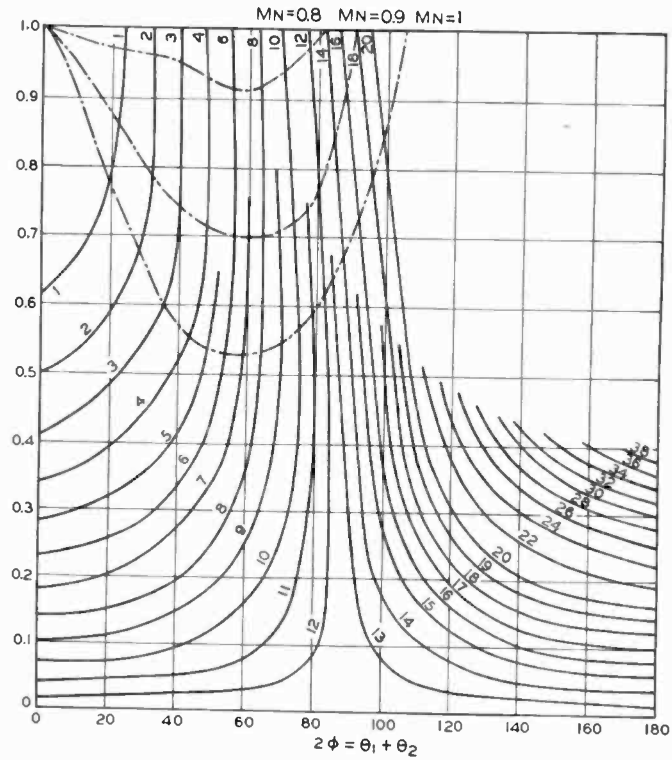


Fig. 7
$M_N = 1.0$, curves labeled values of 100δ.

greater than 20 per cent. It will be seen, as the arguments are developed, that it is very seldom in the practice of asymmetric-side-band broadcasting the curves cannot be used because it is very seldom that M_N is large when 2ϕ is of medium value and ξ greater than the values shown by the error curves. Obviously from (21) if M_N is less than 1, δ is proportionately reduced.

Curves of Constant Values of M_A Plotted Against ξ and ϕ

The value of M_A is given in (22). This expression may be converted to

$$\xi = -\cos 2\phi \pm [\cos^2 2\phi + (4M_A^2 - 1)]^{1/2}. \qquad (34)$$

ξ must always have values lying between 0 and 1. If $M_A = 0.5$ then

$$\xi_{M_A=0.5} = -\cos 2\phi$$

and $\xi_{M_A=0.5}$ lies between the values 0 and 1 when 2ϕ equals 90 and 120 degrees, respectively.

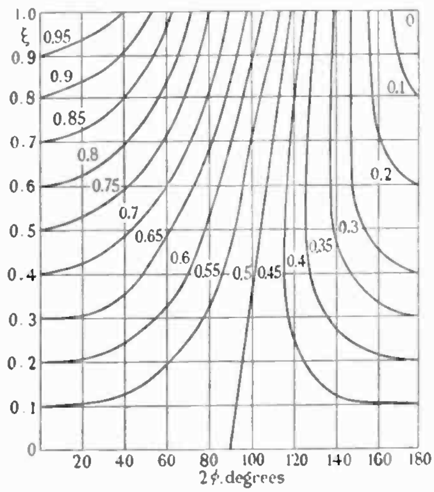


Fig. 8

The implications of the approximation reveal that the value of the fundamental is correct for all values of ξ, ϕ, and M_N. The curves of Fig. 8 derived from (22) and (34) are therefore completely reliable.

Analysis of the Sound Spectrum

The expressions derived in the foregoing section, relating wave-envelope distortion and carrier demodulation with modulation depth and asymmetry of side-band components, must be supplemented, in order that the side-band cutting filter may be correctly designed, by a knowledge of the function relating modulation depth with modulation frequency.

That is to say an analysis must be made of the maximum intensities at different frequencies of the components forming sounds typically transmitted in broadcast practice.

Many such analyses have in fact been made. The author believes that it is sufficient for the purposes of asymmetric-side-band technique to choose a curve, given in Fig. 9, which copies the sensation level curve, as representing a general *probability* for the relative maxima of the modulation components at different audio frequencies. It is assumed that modulation never exceeds 80 per cent. There is 2 decibels difference between 80 and 100 per cent modulation. It is not said that in all cases and at all times the maxima will not exceed the relative values shown in Fig. 9. If the more exotic type of program director chose to broadcast a concerto in which the principal solosit was a bat, accompanied on the Galton whistle, doubtless the curve of Fig. 9 would not be representative. On the other hand, mercifully, neither would the program item!

It seems rather logical to assume that the relative intensities of musical instruments and voice components should follow the sensation-

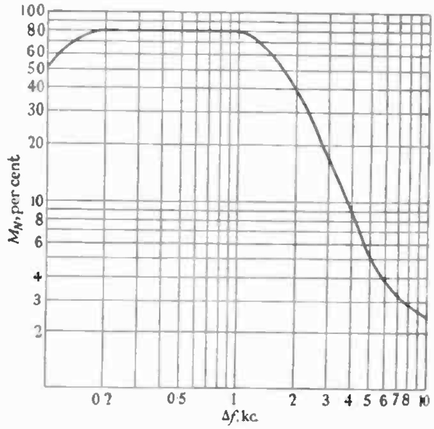


Fig. 9

level curve because, if they did not, the ear would object. The evolution of music has, obviously, been guided by the ear.

If the foundation of these arguments is proved to be unsound and if research and experiment reveal that in fact there are types of program items which, when analyzed, prove that an occasional component rises above the predicted relative level the choice may still be justified by the rarity of the aberration.

Indeed and in sum the soundness of the choice should be judged more in terms of the performance of the asymmetric system than by its comparison with results which seek a complete expression rather than an assessment of probability.

II. DESIGN OF ASYMMETRIC SYSTEMS

General

It is inevitable that distortion should be introduced when part of one side band is cut away and asymmetry between side-band components thereby introduced.

The basic conception of the asymmetric system is that the modulation depth demanded reduces as the frequency of the modulating components is higher and hence, for a constant and negligible distortion, the asymmetry between side-band components and, hence their attenuation, may be greater the more removed the side-band-component frequency from the carrier frequency.

Required Attenuation Constant, ϕ Assumed to be Zero

Let it be assumed that a network for attenuating the outer parts of one side band can be designed which introduces the necessary magnitude asymmetry, i.e., attenuation, without introducing phase asymmetry.

It is seen from (23) that we may write in this case that

$$\delta_\xi = \frac{M_N}{8} \frac{(1-\xi)^2}{1+\xi}.$$

It has also been seen that the carrier is demodulated by the action of the filter, this demodulation increasing as ξ is smaller. Thus at the output terminals of the filter the new modulation factor, when phase asymmetry is absent, is, from (24),

$$M_{A\xi} = \frac{M_N}{2}(1+\xi).$$

At low frequencies of modulation, where both side-band components are substantially equally transmitted, ξ is unity and $M_A = M_N$. At high-modulation frequencies, where the filter has substantially removed one side-band component completely, $M_A = 1/2\ M_N$. Thus the transmitter characteristic, given these conditions, would not be level and so "correction" must be applied.

This correction of the transmitter-frequency characteristic is performed by multiplying the intensity of the input modulation components by $M_N/M_{A\xi}$ or from (24) by $2/(1+\xi)$.

Thus when the transmitter has a level frequency characteristic

$$\delta_{\xi c} = \frac{M_N}{4}\left(\frac{1-\xi}{1+\xi}\right)^2. \tag{35}$$

This may be converted to write

$$\xi_c = \frac{(\sqrt{d_0}-1)^2}{d_0 - 1} \tag{36}$$

where,

$$d_0 = \frac{4\delta_{\xi 0}}{M_N}$$

and $\delta_{\xi 0}$ is a constant value of δ_ξ, the distortion introduced by magnitude asymmetry.

It we assume that the fully transmitted side band suffers practically negligible attenuation then α, the required filter attenuation constant, is given in decibels by

$$\alpha = 20 \log_{10} \frac{1}{\xi_c} \tag{37}$$

or from (36)

$$\alpha = 20 \log_{10} \frac{d_0 - 1}{(\sqrt{d_0} - 1)^2} \tag{38}$$

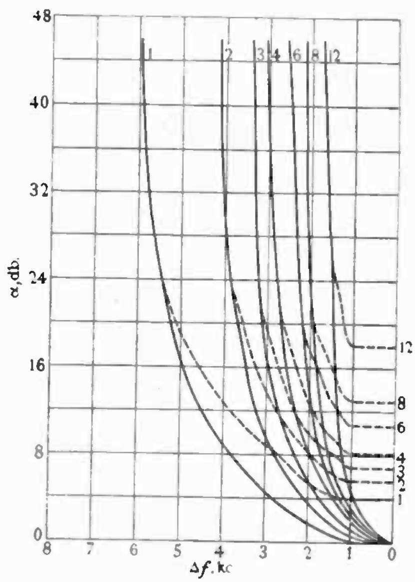


Fig. 10

Criterion attenuation curves, assuming no phase asymmetry, can now be plotted using (38). The value of $\delta_{\xi 0}$ is chosen to be negligibly small. The value of the side-band frequency at which α has different values is given by $f_s = f_0 \pm \Delta f$ the use of the sign depending upon whether the upper- or the lower-frequency side band is attenuated. The value of M_N is given from Fig. 9 for different values of Δf. The frequency scale of Fig. 10 and subsequent figures concerning attenuation and phase characteristic is that of the side-band-to-carrier difference frequency, i.e., the modulation frequency Δf.

It is seen from (38) that as d approaches unity, i.e., when $M_N = 4\delta_{\xi 0}$, that α becomes infinite. The frequency Δf_∞, at which $M_N = 4\delta_{\xi 0}$, is given, assuming a value for $\delta_{\xi 0}$, from Fig. 9. It is also seen that α is finite when $\Delta f = 0$. This occurs because (38) assumes that some constant value of distortion is tolerable at all modulation frequencies. It would be impossible on the one hand to design a filter which fully transmitted components of the carrier frequency and the carrier frequency plus an infinitesimal value of frequency while it attenuated components having the carrier frequency minus an infinitesimal frequency (or vice versa), and it would be foolish to introduce distortion at side-band frequencies, where it is unnecessary and incidentally impossible in practice to do so, and so (38) must be taken only to apply to the higher frequencies where M_N is less than its maximum.

Thus in Fig. (10), showing the criterion attenuation curves for different constant values of distortion, dotted lines are shown for values of frequency where M_N has its maximum values, and the full-line curves join the points where (38) truly applies to the point of zero attenuation at frequency f_0. The full lines represent both the practical and desirable curves.

Choice of Design of Side-Band-Attenuating Networks

The above analysis, deriving criterion curves of attenuation, has assumed that phase asymmetry is not introduced in the filter networks. It remains to design a network which will approach in its performance the criterion attenuation curve without introducing phase asymmetry. It is clear from Figs. 7 and 8 that phase asymmetry, if it exceeded certain values, if ξ were of medium values, would introduce far worse distortion than that chosen as negligible.

Considering only the question of how to get the desired attenuation it is clear that this would be most nearly approached by using a constant-k filter terminated by an f_∞ section.

A study of the characteristics of such a network reveals however that while the required attenuation could be approached the phase characteristic of the combined network, and particularly the effects on that characteristic of the f_∞ section, would introduce very large phase asymmetry when ξ has medium values. This implies that this type of network would introduce intolerable distortion of the wave envelope.

Using only one constant-k section reduces phase asymmetry to negligible values but does not give even a reasonable approach to the requirements of attenuation.

The cascade connection of constant-k sections would give an approach to the required attenuation but introduces, once more, a phase asymmetry sufficient to produce considerably more distortion than would be present if magnitude asymmetry existed alone.

Examination proves that in all these networks the root difficulty is that the flexure of the phase characteristics over the range of fre-

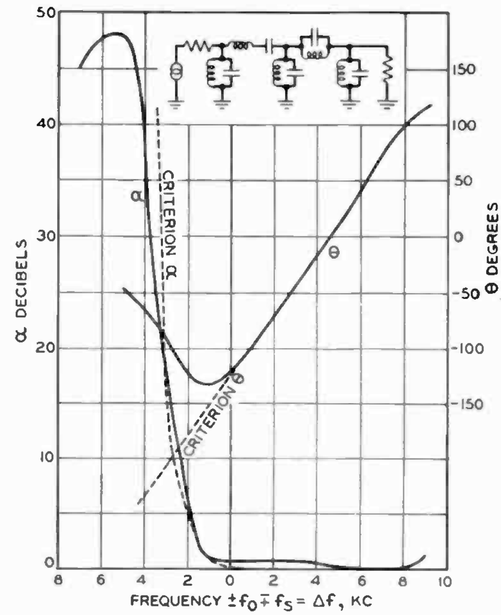


Fig. 11

quencies between f_0 and $f_0 \pm \Delta f_\infty$ is always in the same sense. It appears that any image impedance association of filter sections must reveal this difficulty and that therefore a new approach to the problem must be made.

The solution devised by the author and proved hereafter to be completely successful involves associating the performances of a constant-k band-pass section with that of a constant-k band-eliminator section, the cutoff frequencies being necessarily different. It is obvious from first principles that such networks cannot be directly associated because they would not in such a case be even approximately terminated in their image impedances. Thus it is repeated that it is their separate performances which must combine in the resultant. This implies that a means must be arranged to decouple the two networks, each performing, undisturbed, its separate functions, and it is clear that a practical way to achieve such decoupling is to use valves.

It is perhaps worth while to give a quantitative analysis of the performance of the different networks in order to prove the superiority of the solution to combine the band-pass and band-eliminator filters.

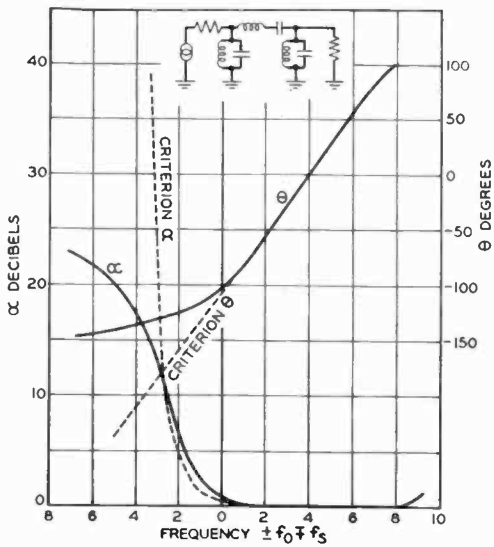


Fig. 12

To this end three separate types of network will be studied, i.e., the constant-k band-pass section terminated by an f_∞ band-pass section (Fig. 11), the constant-k band-pass section alone (Fig. 12), and the constant-k band-pass section in association with the constant-k band-elimination section (Fig. 13).

The method of analysis is first to show in Figs. 11, 12, and 13, the diagram of connections of the network studied and its attenuation and phase characteristics (obtained experimentally). Such characteristics are shown in the diagrams compared with the criterion attenuation and phase characteristics, i.e., those which would give the least possible distortion.

The scale of frequencies given in the diagrams is one representing the difference frequency between carrier and side-band frequency and is therefore a scale of modulating frequency.

Figs. 14, 15, and 16 show the ξ and ϕ characteristics of the networks and also δ and M_A.

The former parameters can of course be obtained from the diagrams giving the attenuation and phase characteristic of the networks and the latter by knowing M_N for the given modulation and hence side-band frequency. Thus at any side-band difference frequency we

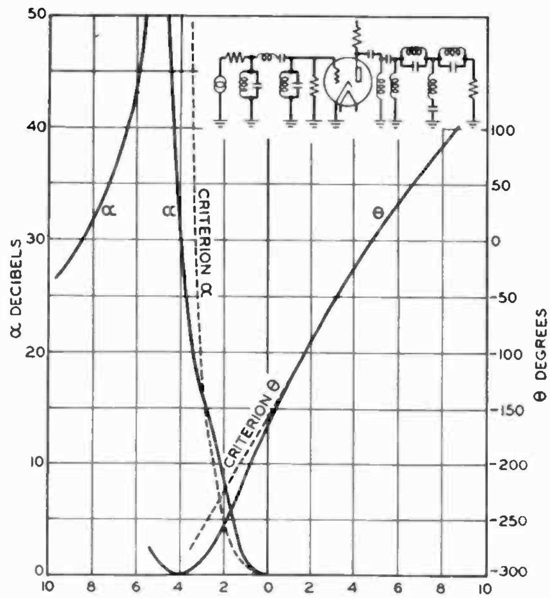


Fig. 13

know M_N (Fig. 9), ξ, and ϕ. It must however be remembered that the value of M_N given in Fig. 9 for any side-band difference frequency and hence modulation frequency must be multiplied by $2/(1+\xi)$ to give its value actually occurring in practice because it must always be assumed that demodulation compensation is necessary. Knowing M_{Nc}, the corrected value of M_N, also ξ and ϕ it is easy from the diagrams of Figs. 7 and 8 to determine the percentage distortion, 100δ, and demodulation M_A, respectively.

The analyses are given in Figs. 14, 15, and 16. It is seen that the constant-k plus the f_∞ section gives a good approximation to the required attenuation but only at the expense of producing large distortion; that the constant-k section alone gives little distortion but fails to achieve as great an attenuation as has been proved to be permissible; and that the association of constant-k band-pass and band-elimination filters gives both the desired attenuation and phase characteristic and so achieves a large attenuation without introducing more distortion than is caused by magnitude asymmetry.

It will be clearer from this analysis why the performance of the network of Fig. 13 is so superior to the others. Thus the cutoff frequencies of the band eliminator are chosen to be a little higher (in the case where the lower frequency side band is cut away) than Δf_∞ where, according to (38), α may be infinite. The phase characteristic of the constant-k band-pass filter is one showing a *decreasing* rate of

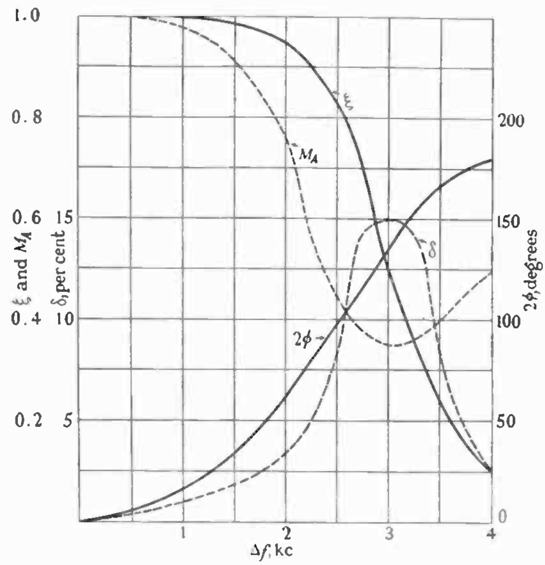


Fig. 14

change over a band of frequencies having limits a little less than f_0 and $f_0 - \Delta f_\infty$. The phase characteristic of the constant-k band eliminator shows an *increasing* rate of change between the limits of this same and vital band of frequencies. In other words, the two phase characteristics have, over this band of frequencies, inverse flexures and their resultant is nearly a straight line. Moreover, at frequencies higher than the carrier frequency, the constant-k band-pass section phase characteristic is nearly a straight line while the rate of change of phase with frequency of the band-eliminator section is so small and, as frequency increases, increasingly tends to be smaller, that it contributes little or no effect, over this pass range of frequencies, to the combined phase-characteristic curve.

Of course very violent phase changes take place at frequencies

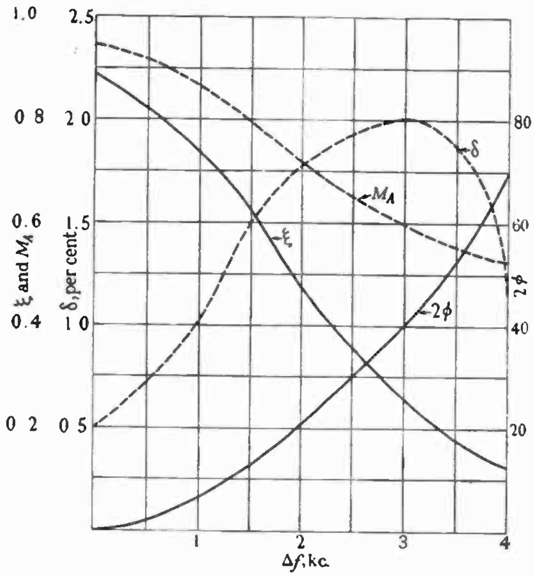

1.0
0.8
0.6
0.4
0.2
2.5
2.0
1.5
1.0
0.5
ξ and M_A
δ, per cent
80
60
40
20
2φ
δ
M_A
ξ
2φ
1
2
3
4
Δf, kc.

Fig. 15

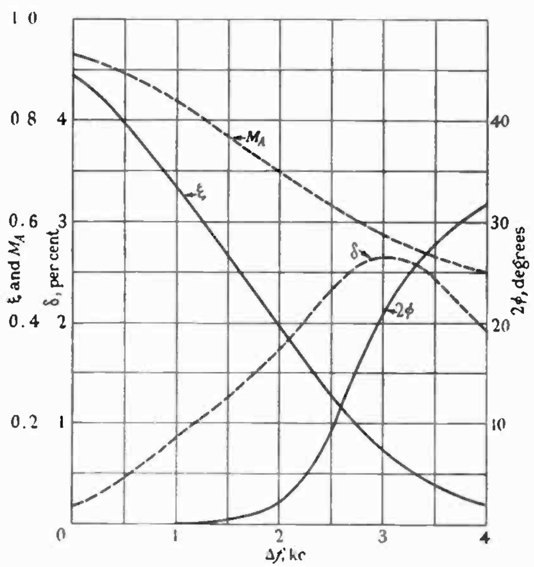

1.0
0.8
0.6
0.4
0.2
0
4
3
2
1
ξ and M_A
δ, per cent
40
30
20
10
2φ, degrees
M_A
ξ
δ
2φ
1
2
3
4

Fig. 16

slightly less than $f_0-\Delta f_\infty$. These phase changes cannot introduce distortion since the attenuation at such frequencies is so large, band pass and band eliminator combining their attenuations, that ξ is virtually zero. Obviously no phase distortion can be present when one side-band component is nonexistent; i.e., when $\xi=0$.

With the other filters studied the addition of further sections of the same type or the addition of f_∞ sections produces phase characteristics which always have the same and even more pronounced flexures as their prototype over the vital range of frequencies f_0 to $f_0-\Delta f_\infty$.

It is concluded therefore that a network wherein constant-k band-pass and constant-k band-elimination filters combine their performances approaches very nearly to the ideal and introduces a very large attenuation of the unwanted side band without introducing more than a negligible distortion.

The scheme has lastly the merit of being extremely simple with the consequent virtues of small cost and ease of adjustment.

Asymmetry of Impedance Characteristics of Asymmetric Filters

A point of importance arises in the practical application of asymmetric-side-band transmission, namely, that care must be taken in associating filters having asymmetrical impedance characteristics with valves, particularly if these are designed to fulfill the functions of modulated magnifiers. Thus evidently if the carrier frequency is located near the cutoff frequency of a filter then the impedance presented to the valves by such a filter is different for a side-band frequency $f_0+\Delta f$ from that presented at the counterpart side-band frequency, namely, $f_0-\Delta f$. Such different impedances may be different not only in value but in character one having the nature of a capacitative and the other of an inductive reactance. Ideally modulated magnifier valves and, to some extent, valves in a class A connection, should work into an impedance having the nature of a constant resistance. The effect of the varying reactance load on the valves is to produce serious distortions not inherent in asymmetric-side-band practice.

A solution to the difficulty is found by making the asymmetric-filter impedance very large compared with the optimum valve impedance, the latter being afforded by a resistance. In order to obtain this condition in practice it is advisable, because otherwise the filter elements have unwieldy values, to use a step-down transformer which is interposed between the valve anodes and the filter, the ratio of the transformer being so chosen that the filter impedance is large compared with the transformed optimum valve impedance. According to this method the primary of the transformer may have a resistance of optimum value connected across it which, in the given conditions, will be practically the sole load seen by the valves. The system is at first sight wasteful inasmuch as the transformer reduces the volts to much smaller values than would be desirable in terms of ideal matching but if the filter is terminated with another transformer which is arranged to step up the volts again and apply these to further amplifying valves the difficulty is overcome.

Fig. 17 shows an arrangement used in certain tests on an asymmetric-transmission system based on the foregoing principles.

Equalizer Design

The carrier is seen to suffer a demodulation at the output terminals of the side-band-cutting filter. In order to level the frequency characteristic the input modulation must be increased above its value in orthodox practice by an amount M_N/M_A.

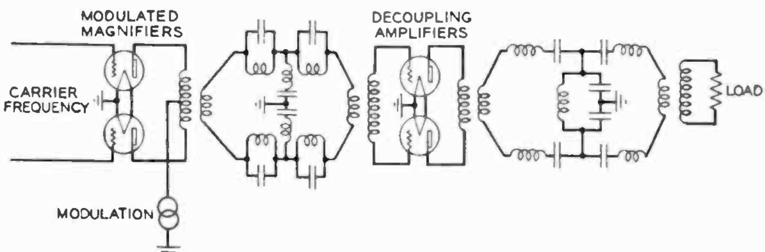


Fig. 17—Schematic of asymmetric transmission system using band-pass and band-elimination filters.

Let $M_N=1$ and let p be the multiplier, then $p=1/M_A$ or $p=1/(2M_A)\times m$ where m is the magnification factor of an amplifier and $1/2M_A$ represents the function of a passive network giving the required frequency discrimination. It has been shown that M_A has boundary values of unity, corresponding to low modulation frequencies (when both side bands are present) and 0.5 at high modulation frequencies when one side-band component is completely removed. It is also seen that, provided phase asymmetry is not greatly pronounced, M_A never falls below 0.5. In this case the conditions for equalization are satisfied if m is constant and equal to 2. In the case where phase asymmetry is pronounced and when M_A falls below 0.5, m must vary with modulation frequency, being equal to 2 at the boundary conditions $M_A=1$, $M_A=0.5$ and greater than 2 when $M_A<0.5$.

It is however unnecessary to deal fully with this latter case because it has been shown that the side-band-cutting filter in which band-pass and band-eliminator filters combine their performances never introduces so much phase asymmetry as to reduce M_A below its critical value of 0.5.

If $\alpha_{\epsilon\xi}$ is the required attenuation constant of the equalizer network and M_ξ the required magnification constant of the amplifier with which it is associated, then the attenuation constant in decibels is

$$\alpha_{\epsilon\xi} = 20 \log_{10} (1 + \xi) \tag{39}$$

and the magnification constant

$$M_\xi = 2. \tag{40}$$

III. THE APPLICATION OF ASYMMETRIC-SIDE-BAND BROADCASTING

Reduction of Interstation Interference

Some five years ago the author published a paper[1] showing a method whereby it was possible to find the minimum frequency separation between the carrier waves of broadcast stations to ensure freedom from interstation interference.

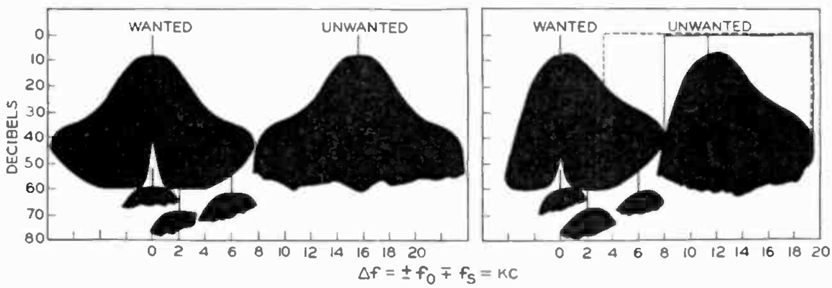


Fig. 18—Minimum separation.

The basic idea, which can be used to give some quantitative assessment of the value of the asymmetric system in economizing channels, was to produce a scaled representation of the maximum limits of the spectra radiated by broadcast stations and maneuver these "shapes" relatively to one another so that in no case should there be any overlapping.

Spectrum Areas

The spectrum of a normal transmission is obtained by drawing the boundaries of the spectrum on a horizontal scale of frequencies (expressed as the carrier frequency minus the side-band-component frequency or vice versa and on a vertical scale of powers).

[1] P. P. Eckersley, "The required minimum frequency separation between carrier waves of broadcast stations," Proc. I.R.E., vol. 21, pp. 193–211; February, (1933).

A vertical line, the top of which touches what is arbitrarily chosen as zero level, represents the carrier. The maximum side-band intensities, for 100 per cent modulation of the carrier, will be virtually 6 decibels below the carrier maximum or zero level. But as there is only 2 decibels difference between 80 and 100 per cent modulation it is common and legitimate practice to attempt to fix 80 per cent modulation as a maximum. Thus in Fig. 18, showing symmetric spectra, the maximum side-band level is 8 decibels below the zero or carrier level.

The level of other side-band components can be derived in their relative value by reference to Fig. 9. The boundaries of the maxima of the side-band components are thus delineated in Fig. 18.

It is helpful in considering the effects of interstation interference to delineate a lower boundary to the spectrum of what we may call the wanted station, i.e., the station desired to be received. If it be assumed that the receiver used is so selective as to give response only between the frequency limits occupied by the wanted station and also that this "perfect" receiver introduces no distortion then the lower boundary of the spectrum is usefully delinated by a curve of the same shape as that giving the threshold of ear hearing. This is clear because in order to find out what level interference must have in relation to the wanted station's intensity to produce interference the boundary must be that above which interference will be audible.

It must thus be assumed that any energy lying between the frequency limits occupied by the wanted station will create *some* corresponding sound-wave energy in the loud speaker, however feeble. The only factor determining whether or not unwanted energy, in the form of parasitical waves or components of the spectrum of an unwanted station, causes audible interference is whether or not it is strong enough to be appreciated by the human ear and therefore the limiting boundary above which intensities are audible and below which they cannot be heard must be the threshold of ear hearing. It is therefore logical to delineate the lower boundaries of the wanted-station spectrum by a line of the shape of this curve.

The only other quantity to be chosen, before the spectrum is enclosed, is the vertical scalar distance at which this line, fixing the ear threshold of hearing, must be fixed relative to the side-band maxima. The determination of this power-level difference involves settling upon what is obviously the contrast level in the reproduction. It is thought that 60 decibels is representative of typical practical conditions and so Fig. 18 is completed as shown.

The preparation of the asymmetric spectra (Fig. 18) is evolved as follows: First it is assumed that no equalizer is used in the transmitter.

In this condition evidently the fully transmitted side band is identical with that created in symmetric modulation. The attenuated side-band maxima are given by subtracting the attenuation constant of the side-band-cutting filter, at given side-band frequencies, from the intensity of the fully transmitted side band at that frequency. But, second, in order to represent the condition when the filter demodulation is compensated for by the equalizer, the intensities of the side bands derived as above must be multiplied by $2/(1+\xi)$.

Fig. 18 shows symmetric and asymmetric spectra in various relative positions, arranged so that no overlap takes place.

Conditions for No Overlap of Spectra

Obviously the conditions for absolute freedom from interference are that the area enclosed by the boundaries of the spectrum shall never be invaded by parasitical unwanted energy. It is not proposed here to consider the effects of the invasions of the types of interference caused by electrical machinery, atmospherics, etc., but it is relevant to consider what conditions must be postulated to be sure that stations on frequency-continguous channels shall not cause interference with a wanted station.

Evidently this can be done, see Fig. 18, by drawing a shape representing the spectrum of a frequency-contiguous or unwanted station and, keeping the lines representing the carrier vertical, sliding this unwanted station shape about that of the wanted, always ensuring that there shall be no overlap. It would of course be fallacious to delimit the lower boundaries of the unwanted-station's spectrum because this extends theoretically infinitely downwards but the higher frequency cutoffs introduced by typical transmission apparatus and by the rapid falling off of the sound intensities as frequency is higher permits the outer boundaries of the unwanted-station's spectrum area to be substantially vertical.

The result of moving one spectrum about the other, always ensuring that no overlap takes place, gives, in proportion to a horizontal scalar distance, the required frequency separation and, according to the vertical scalar difference between the tips of the carrier components, the relative field of the two stations, at the reception location, for the corresponding carrier-wave frequency separation.

Interference Due to Suboptimum Carrier-Wave-Frequency Separations

It is not likely that authority will decide to separate stations by the optimum frequency, revealed as desirable for good-quality reception and freedom from side-band interference, because, if this were done, even though the considerable gains inherent in the adoption of asymmetric technique were used, it would still be necessary to reduce the number of working stations. It is therefore necessary to assess the merits of the asymmetric system in terms of its ability to reduce inter-station interference from the existing amount when, due to symmetric side-band transmission, greater overlap takes place.

The relative degree of interference between symmetric and asymmetric transmission can be assessed as in Fig. 19 by placing the carriers 9 kilocycles apart and studying the relative areas of overlap for different relative field intensities at the reception location.

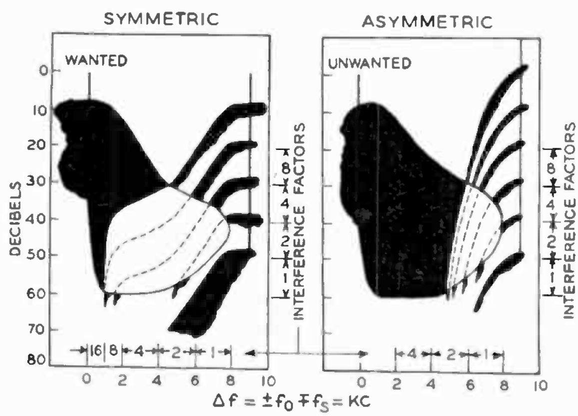


Fig. 19—Interference.

But the ratio of interference factors for the two conditions is not properly judged by a direct comparison of the areas of overlap. Thus some overlap at the outer boundaries of the wanted-station's sacrosanct area is not so deleterious to comfortable reception as if the invasion were nearer to the carrier wave. A user will tolerate some loss of "top" reproduction, if thereby interference can be reduced, or eliminated, but this process of removing the outer parts of the spectrum as interference pushes itself further and further inwards in the spectrum cannot be continued indefinitely, the user, after a certain limit is passed, rather tolerating interference of a certain degree than sacrificing all reasonable fidelity in the reproduction. Thus it is thought logical to give "weight factors" to the ordinates and abscissas of the areas invaded. These are shown in Fig. 19. The weight factors serve to give, in the resulting comparison of interference factors, an appreciation of the fact that invasions nearer to the carrier in frequency and nearer to zero level in intensity are far more annoying than those staged in the more remote parts of the spectrum. This method of assessing the advantages of the asymmetric system moreover seems to be justified in terms of experiments described hereafter.

Summary

The summary of the results obtained from Figs. 18 and 19 is given in curve form in Figs. 20 and 21 showing (a) the required carrier-frequency separation for no interference (symmetric and asymmetric) and (b) the ratio of interference factors between symmetric and asymmetric systems.

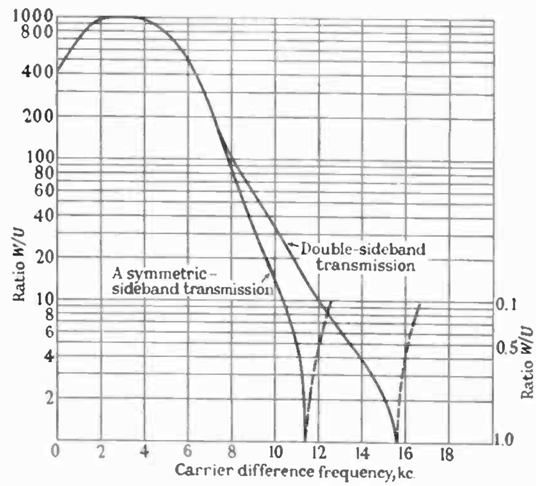


Fig. 20

The results can be summarized as follows: (1) If carrier-wave separations could be increased to the order of 11 kilocycles per second then no interstation interferences would be experienced and high-fidelity reproduction would be probable. (2) If carrier-wave separations were maintained at 9 kilocycles per second, existing interstation interferences would be reduced typically tenfold by the introduction of asymmetric technique. (3) If existing interferences were considered to be in all the circumstances tolerable then carrier-wave separations could be reduced from their present value to 6 kilocycles per second.

Effect Upon Existing Receivers

It is one of the principal advantages of the proposed system of transmission that it could be put into practice without in any way changing the conditions of service to which the listener has become accustomed.

This does not imply that the design and/or adjustment of receivers should remain unaltered if full advantage were to be taken of the more economical system of transmission; it means that the ordinary listener would not, on the introduction of the asymmetric system, remark either a change in volume or an increase in distortion in his received signals.

Thus, so far as volume is concerned, it has been seen that the bulk of the transmitted energy is carried in bands of frequencies close to the carrier and in the asymmetric system double side bands are transmitted over such a band of frequencies. The extra distortion intro-

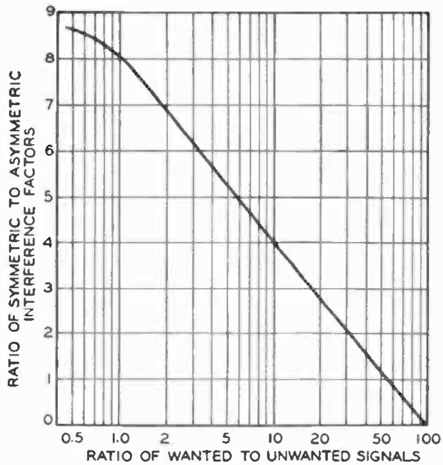


Fig. 21

duced amounting, as has been seen, to an average of $1\frac{1}{2}$ per cent over a band of modulation frequencies between 1500 and 3500 cycles per second, could hardly be remarked by a receiver which, in the generality of cases, introduces per se from $7\frac{1}{2}$ to 15 per cent distortion at full output.

Receiver Tuning

Such postulates however assume that the receiver tuning is adjusted to dispose the response of its filters symmetrically about the carrier frequency. The main object of the asymmetric system is to eliminate the interferences associated with the overlap of side-band spectra. While therefore the application of the method to transmission may prevent such overlap the receiver also must give an asymmetric response before the advantages of the narrowed frequency band of transmission can be appreciated.

Consider Fig. 18. A dotted line is drawn about the unwanted-station's spectrum which for the sake of argument is now considered as the wanted. This dotted line would represent a receiver with its filters symmetrically disposed about the carrier frequency. But obviously in spite of the asymmetric transmission a receiver tuned as shown would experience interference; indeed, because correction multiplies the value of the outer parts of the fully transmitted side band twofold, more interference would be experienced by symmetric reception of asymmetric transmission than if the normal methods of transmission and reception were used. It is therefore necessary for the receiver to be tuned asymmetrically as shown theoretically by the full line in Fig. 18. In sum, the advantages of the asymmetric system can only be realized in terms of a co-operation between receiver and transmitter. This introduces the further point, fundamental to the conception of the system, that the distortion introduced in the resultant, after rectification, is produced not only by the transmission filters but by the combination of all filters, acting upon the modulated carrier wave, which are interposed between the point where the modulated carrier currents are normally generated and the input to the detector.

There are two classifications of reception as local-station and distant-station reception. High-fidelity reception cannot be described as such unless the signal-to-noise ratio is large. Distant-station reception is usually accompanied by noise and the "reacher-out" puts up with a narrowed response to minimize interference and accepts, however grudgingly, and provided the program is interesting, the inevitable interference.

The asymmetric system is hardly necessary for local reception because if the signal-to-noise ratio is large it is unlikely that side-band splash will be pronounced. But in those intermediate circumstances when the field is quite strong enough to give good reception, were it not for side-band interference, then the asymmetric system proves its greatest usefulness. The same applies to distant-station reception except that, when the field is very weak, other interferences often mar reception. The asymmetric technique applied as much to reception as to transmission does obviously minimize noise but not to such a degree as to make it the principal feature of the system.

It is therefore possible to envisage the user of a modern receiver tuning into the local station and getting high-fidelity results by disposing his tuning symmetrically about the carrier thereby introducing no more than the negligible distortion seen to be introduced by the transmission circuits.

At positions and in conditions of reception where side-band splash would otherwise be pronounced the user would tune his receiver asymmetrically thus eliminating or avoiding interferences from stations on frequency-contiguous channels but theoretically at any rate introducing more distortion than is present in the transmission. It is thought, and experiments hereafter described prove, that this distortion is negligible compared with that commonly accepted by the average user as inevitable and it would appear much better to accept this theoretical extra distortion in exchange for the boon of a riddance of the intolerable noise due to side-band splash.

It might indeed be argued that the average listener willingly accepts a degree of harmonic distortion where he resents any interference. Provided the program is interesting harmonic distortion detracts but little from enjoyment, but perhaps the more thrilling the item the more irritating the interference.

Carrier Intensification at the Receiver Detector

There is a solution to the problem of the introduction of distortion in the receiver which would appear to overcome all difficulties and can be represented as a first step in a worth-while evolutionary policy of broadcasting.

It was mentioned in the introduction that distortion of all kinds due to the cutting off of part or all of one side band can be economically overcome by using means in the receiver to intensify the carrier component at the detector without changing the level of the side-band components. Such carrier intensification is in fact a reduction of the value of M_N in the foregoing expressions, and since distortion is shown to be in general proportional to this parameter, its reduction, or conversely the increase of the carrier, results in a proportionate decrease of the distortion.

One might envisage therefore that the introduction of the asymmetric system would encourage receiver manufacturers to apply the carrier intensification systems in their receivers. Such encouragement would be the more forceful if the transmission authorities announced their intention to cut off more and more of the side band, thereby introducing more and more distortion at periods to be agreed upon, say every three years.

The result of such a policy, and assuming the gradual adaptation of receivers to produce the necessary circuits for carrier intensification, would allow us to arrive eventually at the ideal transmission system in which only the carrier and one side band was transmitted.

In the meantime, and even if the required courage to face possible outcry from vested interests proves to be lacking, it is felt that the introduction of the proposed system would be of real public benefit, eliminating or minimizing as it would the all-too-prevalent side-band splash interference without introduction of any dislocation to the existing service.

Adaptation of Transmitters

It is obviously easier to force filter characteristics to conform to a given requirement if the carrier frequency is low and all the experiments so far performed by the author have used carrier frequencies of the order of 30 kilocycles per second. It would be an easy matter to adapt low-power modulated transmitters to the new technique by supplying a modulated-carrier stage, containing the necessary filters and equalizers, the output from which could be applied to some form of frequency-converter stage which, in turn, would feed its output to the first or input stage of the high-frequency amplifier raising the power to the level normally used to energize the aerial.

The problem to adapt transmitters using the high-power system of modulation presents more serious problems. First, it would involve designing the side-band cutting filters to perform their functions at the aerial frequency which in broadcast practice is relatively high compared with 30 kilocycles per second.

Evidently the method to dissociate the band-pass and band-elimination filters from one another could be performed either by resistive networks with consequent loss of power or by valves having a magnification of unity but which would necessarily have to be rather a cumbersome addition to existing plant. It would, perhaps, be worth while introducing a constant-k-filter section only and foregoing the advantages of the greater attenuation introduced by the band-elimination filter.

Experimental Verifications of Theory

Test Transmitter

The British Broadcasting Corporation in the latter part of the year 1936 supplied the author with the necessary funds for the construction of apparatus designed to give a comparison of the quality of reproduction obtainable from the asymmetric system with the best standard obtainable.

The apparatus, a photograph of which is shown in Fig. 22, designed by the author and manufactured by R. M. Radio, Ltd., embodied (a) means for the generation of a carrier of the order 25 to 27 kilocycles per second, (b) means to modulate the carrier at audio frequencies, (c) carrier-frequency filters, one giving symmetric treatment to side bands and the other designed according to the foregoing theory to cut away the outer parts of the lower-frequency side band, (d) a detector and audio-frequency amplifier energizing a high-fidelity loud speaker, and (e) the necessary power supply.

An external hand-operated switch operated relays on the apparatus panels which changed the system from a conventional transmission

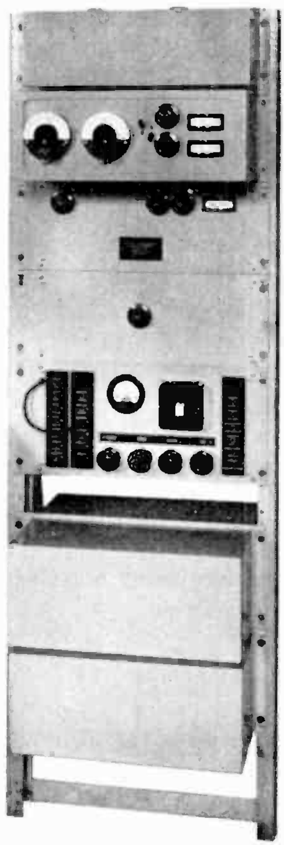


Fig. 22

circuit to the asymmetric type of transmission. The relays thus served to route the modulated high-frequency currents either through a filter in which the carrier was located at f_m or through the constant-k band-pass and band-elimination filters arranged to give, according to the foregoing theory, the least possible phase asymmetry with the maximum attenuation of the outer parts of one side band.

Further relays served appropriately to connect the output terminals of either filter alternatively to the input terminals of a rectifier which in turn fed its output to the audio-frequency amplifier energizing a loud speaker.

It was necessary for a third relay to arrange that the input modulations were routed either through the equalizer necessary, as has been seen, to level the frequency characteristic of the asymmetric transmitter or, when the symmetric system was in use, through a resistive attenuator cutting down the level by half to represent the attenuation introduced, and compensated for by extra amplification, by the equalizer. The introduction of this pad ensured that the modulation level of both transmissions was the same.

All circuits were designed with a large margin of safety in order to ensure that the standard quality should be representative of the best obtainable. After the circuits had been properly adjusted the following performance was obtained from either the symmetric or asymmetric system:

Modulating frequency	500 (cycles per second) to ensure symmetric treatment of side bands in both filters)
Input power to modulators	=0.002 watt (order)
Unmodulated carrier power	=2.0 watts (order)
Power output from audio-frequency amplifier with 90% modulation of carrier	=7.0 watts (order)
Percentage modulation	=90%
Percentage 2nd harmonic measured at loud-speaker terminals	=1.5%
Percentage 3rd harmonic measured at loud-speaker terminals	=negligible
Frequency characteristics	see Fig. 23.

Results of Tests

Tests were arranged so that a comparison of quality could be made. The input modulations were taken from a very high-quality receiver measured to introduce virtually no distortion and tuned to the London regional station transmissions.

The loud speaker was mounted in a quiet room not containing any other apparatus. Red and green lights mounted on the baffle served to distinguish in a given series of change-overs one transmission from the other but switching was arranged so that there was no consistency in the code, the green light (say) applying during one series to asymmetric, in the other to symmetric, but haphazardly. The switch making the change-overs short-circuited the loud speaker during the change over so that the different click sounds introduced by the different relay motions could not indicate a distinction between the two types of transmission.

In sum, everything was done to get a standard of the highest possible quality and all distracting aberrations were removed so that there was no difference between the two reproductions save that inherent in the circuits.

At the time of writing prolonged tests have only been made with the author's staff and it is concluded by them that it is practically impossible, even in terms of a comparison, to distinguish one from the

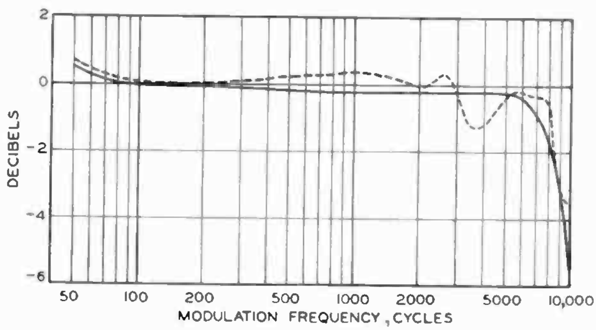


Fig. 23

other. Two observers from the British Broadcasting Corporation were given a short series of tests with the following results:

(1) It is agreed that the system wherein the band-pass and band-eliminator networks are used in combination is a great deal less liable to give audible distortion than that in which an f_∞ section is associated with the constant-k filter. Thus in the previous tests the observers were able to detect the difference in quality between symmetric and asymmetric transmission with some degree of certainty, given certain conditions of modulation and program item. In the tests on the new transmitter the observers were, over a limited period, more often wrong than right in their choice, believing the asymmetric system to be actually more pleasing and less "rough" than the symmetric.

(2) The observers remarked that they thought the quality on both transmissions inferior and that this fact was inclined to vitiate any firm conclusion. The author feels that, while this may conceivably have been due to some fault in the test apparatus or, more probably, in the loud speaker used, nevertheless the very critical analysis of distortion

and frequency characteristics would suggest that in this instance the source itself was not up to standard.

(3) The most remarkable fact adduced was that the observers were inclined to distinguish differences of frequency characteristics between the transmissions; e.g., there appeared to them to be more bass in one than the other. This, in view of the measurements, is very surprising although a quite definitely established fact. The only conclusion one may draw is that a decibel of difference is distinguishable by acute ears provided one supposes that this difference extends over a fairly wide frequency range.

The relevance of the reference to other tests is that similar tests were made before the importance of eliminating phase asymmetry was

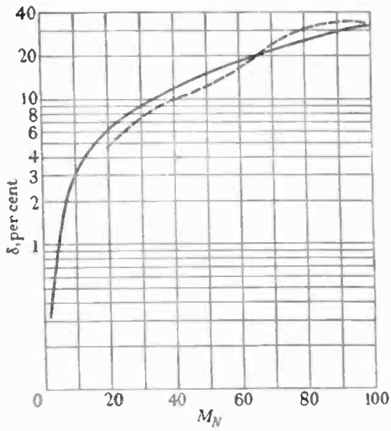


Fig. 24

appreciated. In these previous tests an asymmetric filter was used combining a constant-k and f_∞ band-pass section which has been proved to give violent phase asymmetry and consequent large distortions.

Even with this now provedly undesirable design skilled British Broadcasting Corporation observers were unable always to distinguish the differences between the two transmissions but on balance they were more often right than wrong. Thus taking 100 change-overs as a basic figure, it is estimated that in 50 cases no differences were observed or the observers' opinions canceled one another while of the remaining 50 the observers were right 35 and wrong 15 times.

Tests on Interference

A further experiment was performed in which the output from the transmitter was fed into a frequency converter which in turn supplied the input, at a frequency of about 230 kilocycles per second, to a typical commercial receiver.

An unmodulated carrier was simultaneously introduced into the receiver. This carrier was arranged to have a frequency difference from the multiplied output from the test transmitter of 9 kilocycles per second. The receiver was then tuned to the silent carrier and relative levels of the two carriers, modulated and unmodulated, were adjusted so that the receiver was subject to typical side-band splash interference, the test transmitter transmitting both side bands equally, i.e., representing conditions of symmetric modulation. Moving the switch which changed the test transmitter from a double to an asymmetric side-band trans-

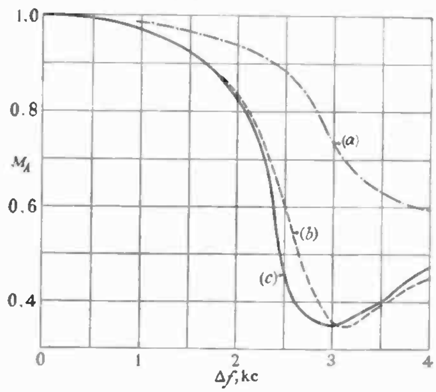


Fig. 25

mitter resulted in a considerable diminution of the interference and it was estimated that, generally speaking, while not completely eliminated the interference, was reduced by at least 20 decibels.

Quality Comparisons with a Commercial Receiver

Opportunity was further taken to make comparisons of quality when the reproducer was a typical commercial receiver and in no conditions of adjustment giving coherent reproduction was any difference noticed between the symmetric and asymmetric transmissions.

Experimental Check of Theory

Lastly, when the transmitter was equipped with the f_∞ section following the constant-k section, which network introduces considerable phase asymmetry, measurements were made to test the accuracy of the theory set out heretofore.

Figs. 24 and 25 show the experimental results compared with those obtained from theory. The agreement is good and proves the practical value of the analysis.

The Koomans System

General

This paper would be incomplete without reference to a system of asymmetric side-band broadcasting developed by N. Koomans of Holland based upon the use of low- and high-pass filters connected in the input modulation circuits of two separate carrier generators the output of which are combined to give the necessary differential attenuation of an asymmetric side-band system.

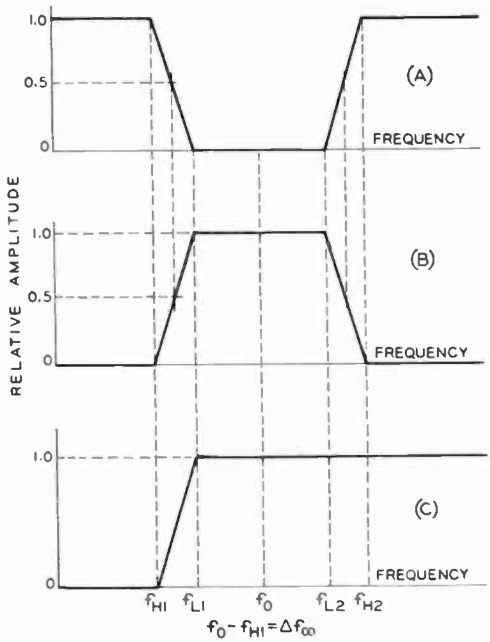


Fig. 26

The author has no greater knowledge of the system than can be gathered from Koomans' British patent specification[2] and offers his apologies in advance to the inventor should any point have been overlooked in what follows.

Principle

The principle of the scheme can be appreciated by reference to Fig. 26.

Fig. 26(A) represents, purely diagrammatically, the side-band-component intensities of a transmitter, modulated from a constant input,

[2] N. Koomans, British Patent No. 450,000.

having a high-pass filter connected in series with the input modulations while Fig. 26(B) is a similar representation of the side-band component intensities of a transmitter having a low-pass filter connected in series with the input modulation circuits.

It is assumed so far as these diagrams are concerned that no phase distortions are present and so the basic principle is treated purely on an amplitude basis.

It is obvious, assuming no phase asymmetry, that if the two transmitters having amplitude-frequency characteristics as shown are combined the resultant will be a symmetric transmission. If however, according to the Koomans system, a side-band-cutting filter is arranged, in the output from the modulated magnifier of the transmitter (the input modulations to which are influenced by the high-pass filter (Fig. 26(A))), which cuts away substantially all of one side band then the combined effect will be that of Fig. 26(C).

It will be clear that over the range of frequencies where the modulation is large, i.e., over a range of frequencies close to the carrier, this side-band-cutting filter cannot introduce distortion because, thanks to the attenuation introduced by the high-pass filter the magnitudes of the side-band components are substantially zero. It is also clear that as regards the other transmitter, the input modulations to which are influenced by the low-pass filter, no magnitude or phase distortion can be introduced since all circuits will present symmetrical characteristics.

Distortion

The distortions introduced will be those due to magnitude and phase asymmetry between counterpart side-band components.

The magnitude asymmetry is given obviously by Fig. 26(C) and ξ becomes less than unity as the frequency of modulation exceeds that necessary to create side-band frequencies less than f_0-f_{L1} and ξ is zero at a frequency of modulation $\Delta f_\infty=f_0-f_{H1}$.

The determination of phase asymmetry is extremely complex and would involve a great deal of experimental and analytical work. Clearly the phase asymmetry cannot be compensated for as in the system previously described.

The root point is however that the use of low- and high-pass filters, according to the technique described, permits or, it would be better to say, forces the rate of change of attenuation with frequency to be extremely rapid. Attenuation must be relatively infinite at a modulation frequency Δf_∞. Taking this as a fundamental then obviously f_∞ sections must be used in the high- and low-pass filters. The rate of change of attenuation with frequency depends upon the ratio f/f_c where f is frequency and f_c the cutoff frequency. The change of frequency between Δf_∞ and some frequency Δf_0 where attenuation must be zero is very small if Δf_∞ and Δf_0 are, as they will be, of the same order as f_c.

In the system formerly described, the rate of change of attenuation of the band filter depends upon f/f_m and f_m has been chosen to be a good deal greater than f_c in the Koomans system. Thus the change in f in the system described formerly is a smaller fraction of f_m than the change of f in the Koomans system and so the rate of change of attenuation with frequency in the Koomans system is relatively much greater.

Evidently phase distortion can only introduce added distortion where ξ has large or medium values and this can only occur where both filters combine their effects over the range of side-band frequencies $f_{H2}-f_{L2}$.

As the flanks of the filter-characteristic curves are made steeper and steeper, so the range of frequencies over which phase asymmetry can be deleterious is less and less until, in the limit, distortions, greater than those introduced inevitably by magnitude distortion, could only occur at one frequency and so by and large we may say that phase distortion in this system, as in the system formerly described, can be neglected.

Comparison of Systems

It is interesting to compare what may perhaps be legitimately described as the Eckersley system of asymmetric side-band transmission with the Koomans system.

If it be agreed that both systems virtually eliminate phase distortion it will be seen that in the Eckersley system this is obtained by the use of a special network acting upon the carrier-frequency currents and in the Koomans system by obtaining an over-all resultant attenuation curve for the cutaway side band which has a relatively much greater rate of change of attenuation with frequency than that given by the Eckersley system.

Distortion

Distortions may be thus compared purely by a consideration of magnitude asymmetry and it is seen that in the Koomans system ξ is unity and distortion zero up to frequencies very close to Δf_∞ whereafter ξ is zero and magnitude distortion is determined by the modulation constant. In the Eckersley system ξ becomes less than unity at modulation frequencies of the order of 1000 cycles per second and gets less and less as modulation frequency is increased becoming zero, as in the Koomans system, at Δf_∞. Thus, considered only from the point of view of distortion, the Koomans system has an advantage over the Eckerlsey system.

Interference Effects

This advantage is not however obtained without a corresponding disadvantage, and that to the author's way of thinking a considerable one, namely, that the Koomans system does not cut away as much of the side band as the Eckersley system.

To make this point clear a diagram is shown in Fig. 27 giving a comparison of the spectra radiated according to whether the one circuit or the other is employed to give the asymmetric effects.

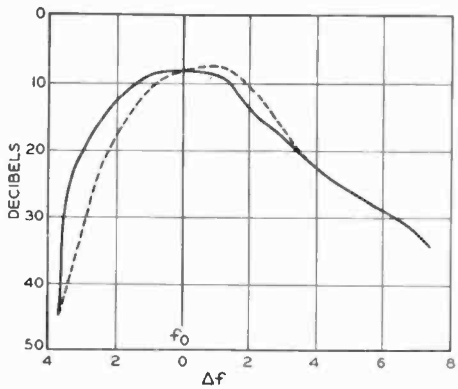


Fig. 27

———spectrum radiated in Koomans system
– – – –spectrum radiated in Eckersley system.

It is unlikely that the carrier waves of broadcast stations will be separated by their optimum whatever asymmetric system is adopted. Overlap is therefore inevitable and it would appear better to cut away as much of the side band as possible provided the resulting extra distortion is not so considerable as to constitute a real disadvantage.

Evidently one cannot "have it both ways" and the inclusion of more of the side band must result in a greater interference albeit less distortion and vice versa.

If it were thought that the reduction of distortion was of paramount importance however, the Eckersley system could be adapted so far as the spectrum shape is concerned to give the same performance as the Koomans system either by simply increasing the number of filter sections of the band-pass and band-elimination type equally or, with fewer sections, using carrier frequencies of the order of 6000 to 10,000 cycles per second.

The point is not fundamental and it is purely a question of what to choose, more side band and less distortion but more interference, or less side band, more distortion, and less interference.

The point that is important perhaps is to appreciate that the Eckersley system is flexible and adaptable according to which policy is chosen, the other relies essentially upon the sharp-cutting filters to reduce phase distortion to negligible quantities.

Receiver Phase-Characteristic Compensation

A point of advantage in the Eckersley system is that the transmitter phase characteristic may be arranged partly to compensate for the receiver phase characteristic.

Thus it is probable that the phase characteristic of the receiver filter has a flexure similar to that given by the constant-k band-pass filter and inverse to that of the constant-k band eliminator.

Fig. 13 shows that with one band-pass and one band-eliminator filter the phase characteristic is dominated to a slight degree by the latter and the addition of another band-eliminator filter would still further overcompensate giving a tendency for correction in the receiver.

Summary

It may be said that the disadvantages of the Koomans system are as follows:

(1) More of the side band is included than is necessary and so a greater interference with stations working on neighboring channels invited.

(2) The correction of the transmitter frequency characteristic becomes more difficult as the attenuation characteristic is steeper.

(3) The adjustment of the high- and low-pass filters to secure the condition of inverse rate of change over the range of frequencies f_{L2} to f_{H2}, Fig. 26, is more difficult as the steepness of the flanks of the characteristics is increased.

(4) The over-all phase characteristic of the received product is a function of the over-all phase characteristic of the transmitter. Ideally phase change should be proportional to frequency. This condition is not obtained in either the Koomans or the Eckersley system but it is suggested that the gentler treatment of phases and magnitudes of the component currents concerned in the latter constitutes a relative disadvantage of the former system. This point is not, it is hoped, put dogmatically and should be considered as purely suggestive.

(5) No phase compensation for receivers can be obtained by the Koomans system.

(6) It is suggested that the circuits involved in the Koomans system are more complex, less flexible, and less easy to adjust than those of the Eckersley system and that the cost of manufacture and adjustment would therefore tend to be higher in the former.

(7) It is necessary to use two modulating systems in the Koomans arrangement. It is true that these only combine their effects over a small range of frequencies and that this range of frequencies is very narrow but it is just at this point where maximum distortion occurs and the introduction of two modulations is not helpful.

The advantages of the Koomans system are as follows:

(1) That less distortion occurs over the range of frequencies 0 to Δf_∞ than in the Eckersley system as described.

(2) The carrier frequency used might be that employed by the station itself although if this were done the filter would have to produce say 20 decibels of attenuation in a change of frequency represented on the average by 3 kilocycles in a million. Assuming this to be practicable the system does not require as does the Eckersley system a frequency converter.

The disadvantages of the Eckersley system are as follows:

(1) The greater distortion introduced if the system as formerly described is used.

(2) It is essential to use frequency conversion.

The advantages of the Eckersley system are as follows:

(1) That it is flexible and able to be designed to give the smaller distortions present in the Koomans system at the expense let it be stressed of including more of the side band. (This virtually cancels disadvantage (1).)

(2) That it has a more uniform phase characteristic.

(3) That it can give receiver phase compensation.

All the points raised are largely theoretical. It is question of policy as to which is better to have less distortion and more side band or vice versa and the Eckersley system could be adapted to either policy.

Perhaps the Eckersley system scores in terms of simplicity, flexibility, and first cost. It is most sincerely hoped in any case that no special pleading has been adopted and that the points are based upon a purely factual analysis.

Conclusion

The analysis, verified by experiment, proves that the outer parts of one side band normally transmitted may be cut away without introducing a more-than-negligible harmonic distortion in the received product.

It is proved that this cutting away of the side-band reduces side-band interference to a tenth of its present value, assuming existing carrier-wave separations remain the same as commonly used today.

It can be shown that the carrier frequencies could be considerably reduced if the interferences as existing today are considered tolerable.

The introduction of the asymmetric system might be the first step in an evolutionary policy wherein more and more of the side band could be cut away as more and more receivers were designed and distributed which, by intensifying the carrier component at the detector, minimized the distortion inherent in the narrowed transmission.

ACKNOWLEDGMENTS

The author wishes to thank the Directors of Messrs. David Allen and Sons, Ltd., for allowing the author both the funds and the time necessary to carry out the analytical and experimental work described in the paper.

The author further wishes to thank the British Broadcasting Corporation for its co-operation both in supplying the funds for building the transmitter and for the interest it has taken in the results obtained.

Thanks are also due to Countess von Zeppelin for her help in devising the approximation and the subsequent derivation of the formula for the demodulation and distortion of the wave envelope when both magnitude and phase asymmetry are present between side-band components.

Mr. R. E. H. Carpenter has been always ready with helpful advice and criticism. Many of the results obtained owe their achievement to the painstaking and careful conduct of the experimental and constructional work involved, by Mr. G. B. Ringham.

Chief Petty Officer Burchell deserves further thanks for his help in the design and construction of the test apparatus.

Bibliography

(1) Societe Francaise Radio-Electrique, French Patent No. 727,097.

(2) P. P. Eckersley, "Asymmetric sideband broadcast transmission," *Jour. I.E.E.*, (London), vol. 77, October, (1935).

(3) O. M. Johnstone and E. E. Wright, "A note on side-band phase distortion," *Wireless Eng.*, vol. 13, p. 534, October, (1936); also G. W. O. H., "Editorial," in same issue p. 517.

(4) H. A. Wheeler and J. K. Johnson, "High-fidelity receivers with expanding selectors," Proc. I.R.E., vol. 23, pp. 594–609; June, (1935).

(5) P. P. Eckersley, "A quantitative study of asymmetric-sideband broadcasting," paper submitted to the Institution of Electrical Engineers (London). January 19, 1938, and read before the Wireless Section March 2, 1938.

(6) P. P. Eckersley, "The required minimum frequency separation between carrier waves of broadcast stations," Proc. I.R.E., vol. 21 pp. 93–211; February, (1933).

Proceedings of the Institute of Radio Engineers
Volume 26, Number 9 *September, 1938*

A CONTRIBUTION TO TUBE AND AMPLIFIER THEORY*

By
W. E. Benham
(Ilford, Essex, England)

Summary—*A formula adapting Maxwell's solution for a plane grating to the plane triode differs from those obtained by previous workers. The triode is simulated by a diode whose electrode separation exceeds the cathode-grid clearance by an amount which is calculated from electrostatic considerations. Corresponding results are derived for the cylindrical triode by means of conformal representation, and are shown to agree better with experiment than the formulas of King and of van der Bijl. The nonuniformity of the electric field due to the presence of the grid, "grating effect," is not marked at the cathode surface, as generally supposed, the field at the cathode being sensibly uniform along the length of the cathode even for grids for which the distance between wires is equal to the cathode-grid clearance. This result is obtained directly from Maxwell's treatise. As the electrons approach the grid the nonuniformity then becomes important and the space-current stream is subject to electron optical considerations. These cause the known departures of g_m from theory in open-mesh tubes.*

The effects of space charge on the measured amplification factor are briefly discussed. It is shown that μ is two stages removed from the Maxwell shielding constant μ_0. The importance of the cathode-plate capacitance is mentioned and an expression given for its approximate value under operating conditions. Llewellyn's calculations for the grid-plate capacitance are discussed in relation to experimental results, which seem to show that this capacitance is greater than the cold value in the negative-grid triode with zero anode load, as predicted by Jarvis, Bakker, de Vries, and North on theoretical grounds.

Variations in μ are automatically included in formulas for current and power output if r_p is defined as the plate resistance measured at the reduced anode voltage $E_p - ri_p$, where r is the resistive component of the load. Maximum output at the fundamental is obtained when the load impedance z_1 is matched to the tube, provided z_1 contains no resistive component which changes as z_1 is adjusted. In case this condition is not fulfilled, or where z_1 is a pure resistance, the optimum load may be determined from the condition $r_p = z_1 + 2z_1(d_{rp}/d_r)$ where r is the resistive component of the load z_1.

Formulas for the first three harmonics produced by a tube obeying an nth power law, the load impedance having in general a frequency characteristic, are given in full. The denominators of the expressions contain $(r_p + z_{n'})$ to the first, third, and fifth degrees, respectively, where $n' = 1$, 2, or 3. (The form of transit-time solutions is later shown to be similar.) Harmonics are described as of minor importance per se as affecting audio-frequency distortion, but it is shown how distortion by intermodulation between strong and weak components of the input signal can amount to 25 per cent of the weaker signal. The main features of Harries' experiments are explained by simple theory, in which an account is given of the great advantages of complex

* Decimal classification: R130×R132. Original manuscript received by the Institute, June 7, 1937. Revised in proof, May 3, 1938.

notation in engineering calculations, with special reference to multiple angles. Conventions are developed for interpreting expressions such as $z(e^{j\omega t})^3$, where z is complex, in terms of the real components of z and sines and cosines of ωt. The results are extended to the resolution of vectors such as $ze^{j\omega_1 t}\cdot e^{j\omega_2 t}\cdot e^{j\Omega t}$ which is an important case in regard to cross modulation.

Following a sketch of the main developments in the transit-time theory of vacuum tubes a new technique for the space-charge-limited diode is explained which depends directly on the simple and universal conception that a charge is propagated without change. The solution is in such a form as to permit evaluation of the nth harmonic from knowledge of $\tau_0 \cdots \tau_{n-1}$ only.

Neglecting space-charge forces, the technique is first illustrated in its application to the first three harmonics in the temperature-limited diode, and the results confirm the correctness of Sloane and James' graphical solution in a special case. In particular, the third-order contribution to the fundamental corrects the phase from +45 degrees to +42 degrees.

The application of the technique to the space-charge-limited diode is illustrated in some detail. For the inclusion of initial velocities an expression is given under general boundary conditions which is the same as already given by Llewellyn in a slightly different form. An extension to large signals including third-order curvature brings out an apparent omission in Llewellyn's expression for the second-order variation time. Equation (91) permits knowledge of the wave form of the total current if that of the input potential is known. Application is made first to the analytically simplest case where the total current is free from harmonics; the potential harmonics are shown on Fig. 3. Sloane and James' graphical solution is less accurate here but represents a creditable achievement. The case where the input potential is free from harmonics is next in order of simplicity. The increase with frequency of the harmonics of the total current may be said to be pronounced, the approximate transit-angle correction factors being $(1+(13/300)\xi^2)e^{+3j\xi/10}$ (fundamental); $(1+(15/200)\xi^2)e^0$ (second harmonic) and $(1+(93/200)\xi^2)e^0$ (third harmonic). These are accurate only for $0<\xi<\pi/4$, for larger values of ξ phase angles differ from zero. The course of these expressions is indicated. Applications to receiving tubes are investigated, an expression being given for the electron convection current at the plane of the control grid of a triode, including the second- and third-order terms. In case the circulation of displacement currents in the input circuit gives rise to appreciable second-order potentials the tube possesses an apparent second-order admittance of which the input conductance may be important though the input capacitance is unchanged from the first-order value. Modulation products at ultra-high frequencies are shown to involve functions of as many transit angles as there are frequencies concerned, and it is pointed out that in many cases rough estimates of their importance may be made from the harmonic analysis given.

In Appendix I following a critique of Pidduck's treatment of the cylindrical magnetron, discussion centers mainly around an expression giving the first-order impedance between a pair of plane electrodes, neither of which need be the actual emitter, under the most general boundary conditions including the presence of an applied magnetic field H tilted at any angle to the plates. The expression reduces to Llewellyn's expression when $H=0$ or when magnetic and electric fields are aligned. By judicious attention to limits the "Seriesschaltung" into which the impedance naturally resolves itself (resistance and capacitance in series) is converted to the equivalent "Parallelschaltung" in the case of low frequencies. The capacitance of the "Seriesschaltung" assumes an infinite value at low frequencies, but the "Parallelschaltung" brings out the existence of negative capacitance for a range of H in the neighborhood

of the critical value. The "Parallelschaltung" may then be shown to be equivalent to another "Seriesschaltung" consisting of a resistance in series with an inductance. Negative resistance arises at low frequencies and the plane magnetron can form part of a self-maintained circuit for natural frequencies of almost any value whatever; the external circuit must be capacitative in order to tune the magnetron, but owing to stray capacitances it is not always possible in practice to stop the magnetron oscillating, for certain values of H. Appendix II explains how calculations for the rectification effect, previously regarded as onerous, may be effected in a few lines in closed form. The published formulas of Benham (1931) and of Llewellyn (1935) are substantiated. Appendix III lists eighteen transit-angle functions, showing their relation to one another and their physical significance, Appendix IV deals with the third-order contribution to the fundamental. In the space-charge-limited diode this contribution to the total current increases with frequency according to the factor $(1+\xi^2/8+\cdots)$ *exp. (0).*

Introduction

THE scope of the present paper will be evident from the summary, though it may be as well to stress the fact that fourth and higher harmonics are not given consideration. The nonlinear theory of electric oscillations[1] lies almost entirely outside the scope of the paper, although it is hoped that the method of approach will be of interest to this field. In particular, the procedure adopted in the present paper is diametrically opposite (not opposed), in that we start from the fundamentals of the tube itself and find out as much as possible from the minimum number of harmonics (first, second, and third), whereas in nonlinear theory it has been the custom to start from circuit phenomena and argue back to the tube, the harmonic coefficients obtained being due, however, not in all cases to tube characteristic curvature but, in addition, to other complications, such as the fact that the tube may only be passing space current for a fraction of the cycle. The analysis of the present paper is exact only in cases where the input voltage never exceeds a value sufficient to cut off the space current for an appreciable fraction of the cycle, but some of the conclusions reached are qualitatively applicable to oscillators and class B and C amplifiers. The main interest of the paper is, however, in regard to tubes operated with class A bias.

Considerations of the influence of finite electron-transit time at high frequencies are deferred until later in the paper. The temperature-limited diode, gives harmonics due entirely to finite transit time. No additional low-frequency transit-time phenomena, such as $C=(3/5)C_0$ and Bakker and de Vries' negative electron damping in the temperature-limited triode, arise from the inclusion of second- and third-degree terms, a careful study of which suggests that, in the space-charge-

[1] Balth. van der Pol, "The nonlinear theory of electric oscillations," Proc. I.R.E., vol. 22, pp. 1051–1086; September, 1934.

limited diode, all harmonics will reduce to those predicted by the static characteristic, at very low frequencies. At high frequencies, however, the influence of transit time on harmonics may be said to be marked, and it is just possible that higher-order harmonics will eventually be shown to be predictable only by the inclusion of finite transit time. If in places I should appear dogmatic, it will be understood that this is for the sake of brevity, in view of the length of the paper.

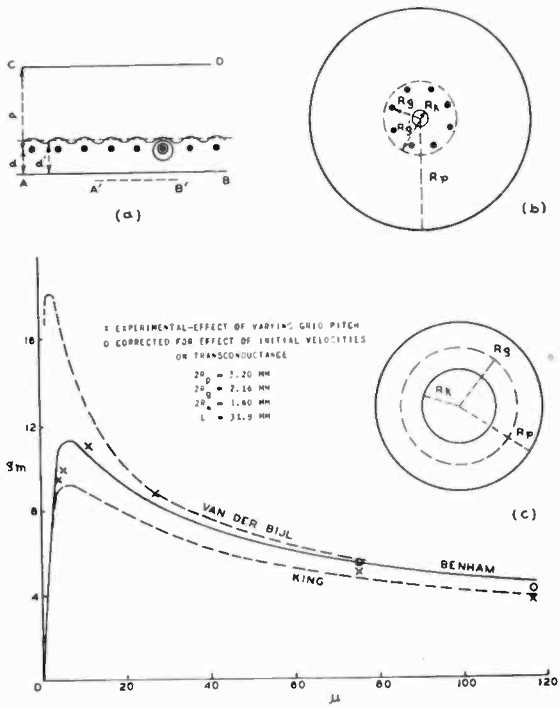


Fig. 1

(a) Position of diode equivalent plane, given by

$$d' = d + \frac{a + d}{\mu_0}$$

where, with n = number of grid wires per unit length,

$$\mu_0 \doteq \frac{2\pi n a}{-\log(2 \sin \pi n c)}.$$

Diagram corresponds to $\mu_0 = 10.5$ nearly. The true equipotential graduates about d', and typical equipotentials in the neighborhood of a grid wire are also shown. The line $A'B'$ has no significance for the plane triode, but see Fig. 1(b).

I. The Electrostatics of a Triode, and the Influence of Space Charge

As is well known, the formulas available for the space current in a triode depend in the first instance on electrostatic considerations. It is probably not so generally realized, however, that equation (12), p. 312, vol. I of Maxwell's treatise gives directly the electrostatic field at the cathode surface of an infinite plane triode due to the combined influence of the anode and grid. All we have to do is to replace Maxwell's symbols as follows (see Fig. 1(a)):

$$\alpha = \frac{a}{\mu_0}; \quad b_1 = d; \quad \frac{b_1 + b_2}{b_1} = \frac{d + a}{d} = \gamma; \quad V = V_g; \quad V_2 = V_p; \quad V = 0$$

and the equation becomes

$$-4\pi\sigma_1 = \frac{\mu_0}{\gamma + \mu_0} \frac{\left(V_g + \frac{V_p}{\mu_0}\right)}{d}. \tag{1}$$

Equation (1) is subject first to Maxwell's assumption that the grid-wire diameter $2c$ is small compared with the distance l between wires. Maxwell's second assumption, that l is small compared with both d and a is of importance in connection with the potential distribution in the neighborhood of the grid wires, but by his equation (5), p. 311, the potential distribution when $y = d$ and when $y = -a$ appears to be so nearly uniform that (1) is thought to be reasonably free from error for values of a and d equal to l. On this basis the geometry indicated

---

(b) The effect of rotating Fig. 1(a) about an axis parallel to AB and distant $a + 0.75$ centimeter from the line CD. The grid wires are then circles about $A'B'$ as axis, but are here shown as sections of longitudinal wires to show spacing.* For longitudinal wires n = number of wires per unit of pitch circle circumference.

μ_0 is increased by this operation from 10.5 to 12.2, while the diode equivalent cylinder is closer to the grid than might have been expected.

$$\log \frac{R_g'}{R_k} = \log \frac{R_g}{R_k} + \frac{1}{\mu_0} \log \frac{R_p}{R_k}$$

$$\mu_0 \doteq \frac{2\pi n R_g \log \frac{R_p}{R_k}}{-\log(2 \sin \pi n c)}.$$

* Van der Pol has shown that a given length of wire of given thickness (whether in the form of a spiral, parallel rods, or mesh) will give rise to the same value of μ_0 when arranged at the grid surface in as uniform manner as possible. See G. J. Elias, Balth. van der Pol, Jr., and B. S. H. Tellegen, *Ann. der Phys.*, vol. 4, pp. 370–406, September, (1925).

in Fig. 1 would correspond to negligible grating effect[2] at the actual cathode surface.

Now (1) may be interpreted in a number of ways, as follows:

(1) The electrostatic field due to the combined action of the grid and anode is reduced in the ratio $\mu_0/(\gamma+\mu_0)$ due to the fact that the grid is not a continuous surface.

(2) The "lumped" potential V_g+V_p/μ_0 acts, not at the grid, but at a plane $d(\gamma+\mu_0)/\mu_0$; i.e., slightly farther out than the grid.

(3) The effective potential at the grid plane (d) is reduced in the ratio $\mu_0/(\gamma+\mu_0)$.

(4) The electrostatic field due to the grid is reduced from (V_g/d) to $(V_g/d)(\mu_0/(\gamma+\mu_0))$; and that due to the anode is reduced from $V_p/(a+d)$ to $V_p/(a+d(1+\mu_0))$ due to the presence of the grid. (Note: This conception leads readily to the relation $\mu_0=C_1/C_0$, used later.)

(5) The potential $(\mu_0V_g+V_p)$ acts, not at the anode, but at a plane given by $a+d(1+\mu_0)$; i.e., at a plane μ_0d farther from the cathode than the anode.

(6) The potential V_g acts, not at the grid, but at a plane $d((\gamma+\mu_0)/\mu_0)$, while the potential V_p must be taken to act at the plane $a+d(1+\mu_0)$.

Now all of the above interpretations are equivalent in the cold triode, but if we are to extend them to cover actual operating conditions the utmost care is necessary in deciding between the various possibilities, and in their application to the evaluation of anode current it is possible to obtain a variety of formulas. Let us denote $d((\gamma+\mu_0)/\mu_0)$ by d', and define d' as "the diode equivalent plane." Then a plane diode having the interelectrode distance d' passes the space-charge-limited current $2.34\times10^{-6}(AV_l^{3/2}/d'^2)$, where V_l may also be taken equal to the "lumped" potential $(V_g+(V_p/\mu_0))$. Then, in accordance with interpretation (2) the anode current of the triode having interelectrode distances d, $d\gamma$ will be the same as in the diode of interelectrode distance d'. On the other hand, if we adopt interpretation (1) or (3) we obtain a result $((\gamma+\mu_0)/\mu_0)^{1/2}$ times as large. There is still the interpretation corresponding to King's solution,[3] namely that the field due

[2] The term "Inselbildung," literally "island formation," rather suggests the phenomenon of patches or islands of activity in the emissive material. For this reason the term "grating effect" is put forward as being more descriptive in English of the nonuniformity of electric field caused by the presence of the first grid, and, to a lesser extent any outer grids as well. The effect is included in Maxwell's treatment of a plane grating if the full expression for the potential distribution is employed (see plate XIII, vol. I, Maxwell's treatise).

[3] R. W. King, "Calculation of the constants of the three-element vacuum tube," *Phys. Rev.*, vol. 15, pp. 256–268; April, (1920). In the notation here used

to the potential $\mu_0 V_g + V_p$ is reduced in the ratio $\mu_0 : \gamma + \mu_0$, but King then takes this potential to act at a plane given by $(a+d)$ which seems to me to be unjustified (Compare (5)). At all events, King's formula is analytically unsatisfactory, since when μ is made large in it, we do not get the correct limiting formula for a diode coinciding with the cathode-grid space. On the other hand van der Bijl's formula[4] fails at low values of μ. Interpretations (4) and (6) would be ideal if a simple way could be found for adapting them to the space-charge equations. Failing this, the best that can be done is to refer the action of the triode to that of the equivalent diode in accordance with interpretation (2). Electron-optical effects—grating effect in the neighborhood of the grid wires—will sort the anode current stream into pencils but the effect of this on the actual value of the anode current will be ignored. We then obtain

$$I_p = \frac{2.34 \times 10^{-6}}{d'^2}\left(V_g + \frac{V_p}{\bar{\mu}_0}\right)^{3/2} \tag{2}$$

where,

$$d' = d + \frac{a+d}{\bar{\mu}_0}$$

$\bar{\mu}_0$ = value assumed by μ_0 when space charge is present.

It will be seen in Section II that it is an approximation to take $\mu(= -dV_p/dV_g \mid I_p) = \bar{\mu}_0$, but for the present purposes this difference will be ignored, so that (2) becomes

$$I_p = \frac{2.34 \times 10^{-6}}{d'^2}\left(V_g + \frac{V_p}{\mu}\right)^{3/2}. \tag{2a}$$

For the cylindrical case, we may pass by means of a suitable transformation from (2a), obtaining for a structure of length L,

$$I_p = \frac{14.70 \times 10^{-6} L}{\beta_{g'}^2 R_{g'}}\left(V_g + \frac{V_p}{\mu}\right)^{3/2} \tag{3}$$

King's formula for cylinders is

$I_p = 14.70 \times 10^{-6} L/\beta_p^2 R_p[(R_p - R_k)(V_p + \mu V_g)/R_p - R_k + (\mu+1)(R_g - R_k)]^{3/2}$.

In the case of $V_g > 0$ and $\mu \rightarrow \infty$ we should obtain a diode formula corresponding to V_g applied at grid, as given by (3a) of the present paper.

[4] H. J. van der Bijl's formula is taken as

$14.70 \times 10^{-6} L/\beta_p^2 R_g[V_p + \mu(V_g + \epsilon)/1 + \mu]^{3/2}$.

While this formula was attributed to van der Bijl by the author of reference 5, I have not been able to find it in this form in the references given. I think it was intended to convey that van der Bijl was the first to point out the importance of the quantity $V_p + \mu V_g$, in his paper, *Verh. den Deutschen Physik. Ges.*, vol. 15, p. 338; April, (1913).

where $\beta_{g'}$, $R_{g'}$ refer to the "diode equivalent cylinder" given by

$$\log \frac{R_{g'}}{R_K} = \log \frac{R_g}{R_K} + \frac{1}{\mu} \log \frac{R_p}{R_K} \tag{4}$$

where R_p, R_g, R_K are radii of anode, grid, and cathode, respectively. Equation (4) is the same as equation (141), Chapter II of W. G. Dow's "The Fundamentals of Engineering Electronics," John Wiley and Sons and Chapman and Hall, Ltd., (1937). Readers are referred to this work, which appeared after the present paper was written, for the details in mathematics of the electrostatics of triodes.

Equation (3) differs, however, from Dow's space-current formula (equation (263) of Chapter V). After careful consideration of Dow's arguments I find no reason to change from my own interpretation. Anode currents predicted by (3) agree well with experiment, but Dow's formula appears to give anode currents somewhat too high.

From (2a) and (3) the corresponding transconductances are obtained. Again, as will be seen later, an approximation is involved in taking $\bar{\mu}_0$ or μ to be constant during the differentiation of I_p with respect to V_g. However, such change will be ignored for the present and transconductances calculated from the formulas

$$g_m = \frac{2.34 \times 10^{-6} A}{d'^2} \cdot \frac{3}{2} \left(V_g + \frac{V_p}{\mu} \right)^{1/2} \quad \text{(planes)} \tag{2b}$$

$$g_m = \frac{14.70 \times 10^{-6} L}{\beta_{g'}^2 R_{g'}} \cdot \frac{3}{2} \left(V_g + \frac{V_p}{\mu} \right)^{1/2} \quad \text{(cylinders)}. \tag{3a}$$

For the purposes of comparison with experiment it is more convenient to work with transconductances since anode currents vary so rapidly with grid pitch. On Fig. 1(c) will be found the theoretical curves corresponding to van der Bijl's and King's formulas for cylinders, together with that obtained from (3a). It must be emphasized that the form of the g_m—μ curve varies with geometry, the actual structure on which the curves are based being shown. These curves are not based on any formula for μ, but give g_m as a function of μ using (3a). In Section II it is shown that measured μ is at least two stages removed from the electrostatic μ_0. The experimental points (measured g_m and *measured* μ), also obtained from tubes having this geometry, are corrected for contact potentials and initial velocities, while 15 per cent has been added to allow for the effect of the grid support. The points correspond to averages over a number of tubes. The agreement is, on the whole, best with Benham's formula, though the high μ values lie nearly on King's curve. On the other hand, these points could be expected to suffer, for while the effect of emission velocities have been taken care of experimentally by measuring at the grid bias at which grid current starts ($I_g=1$ microampere), the theoretical curves (corresponding to $V_g=0$, $V_p=100$) were based on initial velocities of zero. On Langmuir's theory[5] the correction required to the experimental g_m values is roughly $(V_l/(V_l-V_0))^{1/2}$ where $V_l=V_g+(V_p/\mu)$ and V_0 represents the effect of emission velocities expressed in volts. Taking $V_0=0.2$ volt the correction factor is 1.085 in the case where $\mu=75$ and 1.14 in the case where $\mu=118$. The points X are therefore shifted to the points O. In other cases the corrections would be very much less and have not been shown.

It should be mentioned that such good agreement with theory could not have been obtained had precautions not been taken to avoid mechanical variations[6] in the tubes. These particular samples were constructed with the sole purpose of establishing agreement between theory and experiment in the case of tubes with small cathode-grid clearance. These designs were available in England in the year 1930 and the tubes known to be particularly free from losses on ultra-short waves. The choice of a rather small anode-grid clearance was governed partly by the consideration that low μ values could thereby be obtained without grating effect (Inselbildung) being serious.[2] Thus, for $\mu=4$ the number of turns of 0.003-inch-diameter wire was no less than 40 turns per inch. It does appear that this effect is tending to upset the theory at low values of μ, but the shift in the optimum value of g_m to the right of the theoretical curve is just what would be expected from a rough calculation of grating effect based on simplified electron-optical considerations. Thus electron optics plays an important part in the cathode-grid space, for open grids, in determining the value of g_m, but in the present case, as already indicated, the value of μ for which these effects begin to come in is very much reduced. As is well known the grid-wire diameter affects the value of μ (see (4)) and if a reduction of grid-wire diameter, rather than of n or of plate diameter, is used in order to effect a decrease in μ, better results both on g_m and on the shape of the anode characteristic are obtained, in general. From the point of view of variable-μ tubes, however, the considerations which apply are special and it seems that the usual practice of removing

[5] I. Langmuir, "The effect of space charge and initial velocities on the potential distribution and thermionic currents between parallel plane electrodes," *Phys. Rev.*, vol. 21, p. 419; April, (1923).

[6] M. Benjamin, C. W. Cosgrove, and G. W. Warren, "Modern receiving valves, design and manufacture, *J.I.E.E.* (London), vol. 804, 84, pp. 401–439; April, (1937). See p. 410.

a turn or of using variable grid pitch, though rather unsatisfactory from the manufacturing standpoint produces a close approximation to the circuit requirements. However, manufacturing problems are outside the scope of the present paper, except that the attention paid to fundamentals may in some small measure help in devising means for improved designs.

II. The Difference Between μ, $\bar{\mu}_0$, and μ_0

For the cold triode the following relations are exact

$$\mu_0 = \frac{C_1}{C_0} \tag{5}$$

$$\gamma = 1 + \frac{C_1}{C_2}$$

whence,

$$\frac{\mu_0}{\mu_0 + \gamma} = \frac{C_1 C_2}{\sum C_1 C_2} \tag{6}$$

where,

$$\sum C_1 C_2 \equiv C_0 C_1 + C_1 C_2 + C_2 C_0.$$

C_0 = cathode-anode capacitance, C_1 = cathode-grid capacitance, C_2 = grid-anode capacitance.

In order to investigate the effects of space charge, we may proceed by *defining* the "hot shielding ratio," as follows:

$$\bar{\mu}_0 = \frac{\overline{C}_1}{\overline{C}_0}\,. \tag{5a}$$

Likewise, we define

$$\bar{\gamma} = 1 + \frac{\overline{C}_1}{\overline{C}_2}$$

while,

$$\frac{\bar{\mu}_0}{\bar{\mu}_0 + \bar{\gamma}} = \frac{\overline{C}_1 \overline{C}_2}{\sum \overline{C}_1 \overline{C}_2}\,. \tag{6a}$$

Now in our application to the space-current formula (2) the difference between the values of d' as calculated using (6) and (6a) may be disregarded as a first approximation, since the factor $\bar{\mu}_0/(\bar{\mu}_0+\bar{\gamma})$ remains in the neighborhood of unity except for low values of $\bar{\mu}_0$.

We must, however, use the relation (5a) in the bracketed factor of (2) so that when changes are made in V_p and V_g, keeping I_p constant, changes in $\bar{C}_1$ and $\bar{C}_0$ must be taken into account. In this way we obtain for the amplication factor $\mu(=-dV_p/dV_g|I_p)$

$$\mu = \frac{\dfrac{\bar{C}_1}{\bar{C}_0} - \dfrac{V_p}{\bar{C}_1}\dfrac{\partial \bar{C}_1}{\partial V_g} + \dfrac{V_p}{\bar{C}_0}\dfrac{\partial \bar{C}_0}{\partial V_g}}{1 - \dfrac{V_p}{\bar{C}_1}\dfrac{\partial \bar{C}_1}{\partial V_p} + \dfrac{V_p}{\bar{C}_0}\dfrac{\partial \bar{C}_0}{\partial V_p}} = \frac{\bar{\mu}_0 - V_p \dfrac{\partial}{\partial V_g} \log \dfrac{\bar{\mu}_0}{\mu_0}}{1 - V_p \dfrac{\partial}{\partial V_p} \log \dfrac{\bar{\mu}_0}{\mu_0}} \tag{7}$$

where μ_0 is the value of $\bar{\mu}_0$ when conditions are such as to permit satisfaction of the equation $\mu = \bar{\mu}_0$; i.e., μ_0 is the electrostatic-shielding ratio given by (4) if C is not too large compared with $1/n$. Actually, of course, the condition $\mu = \bar{\mu}_0$ would be met wherever

$$\left(\frac{\partial}{\partial V_g} - \bar{\mu}_0 \frac{\partial}{\partial V_p}\right) \log \frac{\bar{\mu}_0}{\mu_0} = 0. \tag{7a}$$

Equation (7a) may readily be checked experimentally and it will be found not to hold in general. But in this case $\bar{\mu}_0 \neq \mu$ and we have to fall back on a knowledge of the various interelectrode capacitances and the changes of these with the electrode voltages or space currents. Equation (7a) holds at cutoff, so that it may be said that $\mu = \bar{\mu}_0 = \mu_0$ at cutoff, and only at cutoff.

The above analysis suffices to indicate that the measured amplification factor μ is two stages removed from the Maxwell shielding constant μ_0, so that until further information is obtained, particularly in regard to the small capacitance $\bar{C}_0$, it is unreasonable to look for good agreement between experiment and the electrostatic formulas for μ_0. In some experiments carried out many years ago the writer found by a low-frequency-bridge method that both $\bar{C}_2$ and $\bar{C}_1$ were increased[7] by the presence of space charge, but although an attempt was made to detect changes in $\bar{C}_0$, the cold value C_0 was too small to enable conclusions to be drawn. By Fig. 10 of reference 9 we should expect the effective capacitance between cathode and plate to be less than the cold value due to the presence of the inductance L in series with the resistance R. Now at low frequencies R is equal to the plate resist-

[7] Since this paper was written a grid-anode capacitance less than the cold value (anode-circuit impedance zero) has been claimed by T. Iorwerth Jones, "Inter-electrode capacitances of valves," *Jour. I.E.E.* (London), vol. 81, pp. 658–666; November, (1937). In my discussion to Jones' paper, *Jour. I.E.E.* (London), vol. 82, pp. 220–222; February, (1938), I showed that this may have been due to a method of measurement in which the grid electrode was at radio-frequency earth potential. On page 222 I mentioned the possibility of a negative value of δC_{af} (in our present notation this means $\bar{C}_0 < C_0$). By reference to equation (7b) this would appear unlikely, but it must be remembered that this equation is subject to inaccuracy arising out of the difference between μ and $\bar{\mu}_0$.

ance r_p. If R_0 denotes the resistance of the diode corresponding with the cathode and grid of the triode, (2) may be used to give

$$R_0 = \frac{r_p}{\mu}\left(\frac{\mu}{\gamma + \mu}\right)^2$$

in which the difference between μ and $\bar{\mu}_0$ is ignored. The inductance L then has the value

$$L = L_1 + L_2 \doteq \frac{11}{15} C_1 \frac{r_p^2 \mu^2}{(\gamma + \mu)^4} + r_p\left[1 - \frac{\mu}{(\gamma + \mu)^2}\right]\tau_2$$

where τ_2 is the transit time between grid and plate. The capacitance existing between cathode and plate may then be estimated as follows: Neglecting electron inertia, the cathode appears to be only three quarters its real distance from the first grid in the ideal Langmuir case,[8] or, more strictly, three quarters its real distance from the diode equivalent plane; i.e., $[1-(\gamma+\mu)/4\mu]$ times its real distance d from the first grid. The "true" capacitance between the first grid and the cathode is then $4\mu C_1/(3\mu-\gamma)$ and the "true" capacitance[9] between anode and cathode is given by $4C_1/(3\mu-\gamma)$, bearing in mind (5a). The "effective" capacitance[9] between cathode and anode is then

$$\overline{C}_0 = \left[\frac{4\mu}{3\mu - \gamma} - \frac{11\mu^3}{15(\mu + \gamma)^4}\right] C_0 - \frac{\tau_2}{r_p}\left[1 - \frac{\mu}{(\gamma + \mu)^2}\right]. \qquad (7b)$$

For μ large and τ_2/τ_1 small the above corresponds very nearly to $\overline{C}_0 = (4/3)C_0$, in agreement with Llewellyn's calculations,[10] but it must be mentioned here that I have been unable to adhere to Llewellyn's results[10,11] for $\overline{C}_2$. Some years ago I conducted a series of experiments on triode capacitances and found the grid-anode capacitance always greater than the cold value. Moreover, the calculations of Jarvis, Bakker and de Vries, and of North all give a grid-anode capacitance greater than the cold value, in the case of zero anode load. The possibility of a capacitance less than the cold value appears to me to be

[8] Balth. van der Pol, Jr., "Über Elektronenbahnen in Trioden," *Jahr. T. T. (Zeit. für Hochfrequenz.)*, vol. 25, pp. 121–131; May, (1925). Also "Über Elektronenheweginen in Trioden," *Physica*, vol. 3, pp. 253–275; September, (1923). For negative-grid triodes (space-charge limited) the calculation may be made without reference to electron-transit time.

[9] W. E. Benham, "Theory of the internal action of thermionic systems," *Phil. Mag.*, vol. 11, p. 457; February, (1931). Part II of reference 24.

[10] F. B. Llewellyn, "Equivalent networks of negative-grid vacuum tubes at ultra-high frequencies," *Bell Sys. Tech. Jour.*, vol. 15, pp. 575–586; October, (1936).

[11] F. B. Llewellyn, "Operation of ultra-high-frequency vacuum tubes," *Bell. Sys. Tech. Jour.*, vol. 14, pp. 632–665; October, (1935).

restricted to cases where an anode circuit impedance is present. The capacitance and resistance of the grid-anode space appear to be functions of the load impedance in general, and it is difficult to accept the grid-anode part of Llewellyn's delta network[10] for this reason alone. It may be that I have misunderstood Llewellyn's choice of internal generators.

III. The Effect of Variable μ on Amplifier Output

In cases where μ variation is of importance it naturally makes a difference if we alter the form of the anode-current equation from

$$I_p = K\left(\frac{V_p}{\bar{\mu}_0} + V_g\right)^n \tag{8}$$

to

$$I_p = K'(V_p + \bar{\mu}_0 V_g)^n \tag{8a}$$

where K, K' are constants. Reasons have already been given in Section I for preferring form (8) rather than form (8a). The factor K is still slightly dependent on $\bar{\mu}_0$, since it contains the factor $1/d'^2 = \bar{\mu}_0^2/(d^2(\bar{\gamma}+\bar{\mu}_0)^2)$, but for most cases of practical interest the variations of $\bar{\mu}_0/(\bar{\gamma}+\bar{\mu}_0)$ with μ are unimportant compared with the variations of $((V_p/\bar{\mu}_0)+V_g)^n$ with $\bar{\mu}_0$, and will be neglected in the sequel. While n would generally be taken as 3/2 there is no need to impose this restriction and the sequel will apply for any positive value of n and for grid swings ranging from zero up to the value $(V_g+(V_p/\bar{\mu}_0))$, being the lumped voltage acting at the diode equivalent plane. Actually, we are not restricted to a plane or cylindrical geometry, so that the word "plane" may be replaced by "surface."

We shall begin by considering the effect of making direct-current changes in the values of V_p, V_g, allowing for $\bar{\mu}_0$ variations. It is desirable to regard the load in the anode circuit as presenting impedance $z_{n'}$ to the n'th harmonic,[12] so that z_0 will be the direct-current resistance r. Then in the direct-current condition

$$I_p = K\left(\frac{V_p}{\bar{\mu}_0} + V_g\right)^n = K\left(\frac{E_p - rI_p}{\bar{\mu}_0} + E_g\right)^n. \tag{9}$$

If we adopt the definition "r_p is the tube resistance measured without r in circuit but at the reduced anode voltage $E_p' = E_p - rI_p$", we obtain from (9)

[12] In Sections III and IV n' denotes the order of the harmonic, n having been used for the index of the power law. Strictly speaking $z_{n'}$ is the impedance to n'th-order currents. This will be clear from Section IV, equation (15a).

$$\frac{1}{r_p} = nK\left(\frac{E_p'}{\bar{\mu}_0} + E_g\right)^{n-1}\left(\frac{1}{\bar{\mu}_0} - \frac{E_p'}{\bar{\mu}_0^2}\frac{\partial \bar{\mu}_0}{\partial E_p'}\right) \tag{10}$$

while, in a similar fashion to (7),

$$\mu = \frac{\bar{\mu}_0 - E_p' \dfrac{\partial}{\partial E_g} \log \dfrac{\bar{\mu}_0}{\mu_0}}{1 - E_p' \dfrac{\partial}{\partial E_p'} \log \dfrac{\bar{\mu}_0}{\mu_0}} \tag{11}$$

we obtain under all circumstances

$$i_1 = \frac{\mu v_g}{r_p + z_1}. \tag{12}$$

Thus, in the sequel, all further reference to μ variations will be rendered unnecessary if we adopt the definitions (10) and (11) for r_p and μ, respectively. What this amounts to is, that in the event of direct-current resistance r being present in the plate circuit of the tube, we refer all calculations to the reduced plate voltage $(E_p - rI_p)$, and also that, when this is done, we may regard all μ variations as included in our equations if we proceed without taking these into consideration, but in our final answer these may be included explicitly, if desired, by means of (11).

Since (12) now gives the output of fundamental current, we find that for $r=0$ maximum power output is obtained when $r_p = z_1$, since in this case no change takes place in r_p when z_1 is varied. In the case where $z_{n'} = r$ for all values of n' (pure resistance load), the condition $r_p = r$ no longer holds as the condition for maximum output; we can no longer make changes in the load without affecting r_p. Thus if P = power output

$$P = z_1 i_1^2; \quad \frac{dP}{dz_1} = i_1^2 + 2i_1 z_1 \frac{di_1}{dz_1} = i_1^2\left[1 - \frac{2z_1 + 2z_1 \dfrac{dr_p}{dz_1}}{(r_p + z_1)^2}\right]$$

which is zero when $r_p = z_1 + 2z_1(dr_p/dz_1)$. When the resistive component of z_1 is zero $dr_p/dz_1 = 0$, otherwise $r_p > z_1$ for maximum power output.[13]

[13] It will be understood here that these conclusions refer to the maximum fundamental output obtainable assuming a given plate-supply voltage. This is sometimes of greater practical interest than the case where the plate potential is supposed to be maintained by raising the plate-supply voltage by an amount sufficient to compensate for the drop in r. In this last case the condition $r_p = z_1$ is always found.

While it is believed that these last results are known in informed circles, I have never seen them clearly expressed; moreover they have not, to my knowledge, previously been established in the case of μ variable.

IV. Output on Second and Third Harmonic

The above considerations referred, of course, to the optimum output of fundamental regardless of harmonic distortion. The following analysis will be restricted to fundamental, second, and third harmonics, as all essential features necessary to a proper understanding of cross modulation, intermodulation, and side-tone distortion may be obtained if the third harmonic is included, and higher harmonics than the third will usually not be of importance in properly designed tubes. While it is here tacitly assumed that feedback (anode to first grid) is negligible, the contents of this section are intended to apply in case the load possesses any desired frequency characteristic. In other words, the frequency may be of any desired value such that the internal impedance of the tube may be taken as purely resistive to a good approximation.

In most treatments of harmonic production the dynamic characteristic of the tube is used as a starting point, i.e.,

$$i = a_0 + a_1 v + a_2 v^2 + a_3 v^3 + \cdots \tag{13}$$

where a's may be obtained from dynamic tube characteristics, v being the grid swing measured from the operating point, as usual. Now while this procedure leads to simpler formulas than the one which will be adopted here, it is thought worth while to relate the analysis more intimately to the tube by working from (8). This procedure has also the advantage that estimates may be made without reference to dynamic characteristics. Before proceeding, however, it is worth while to draw attention to the fact that if a method of measuring dynamic transconductance is available it is preferable to start, not with (13), but with the equation obtained by differentiation with respect to v,

$$di/dv = a_1 + 2a_2 v + 3a_3 v^2 + \cdots$$

and to make a series of "dynamic g_m" curves. From these it is a simple matter to obtain the second and third harmonic, the former being proportional to the slope of the g_m curve and the latter to the rate of change of slope (rough estimates may be made by drawing chords on the g_m curves, the third harmonic being proportional to the maximum distance between chord and curve).

A point that arises if we start from (8) is that this equation does not represent the pentode characteristic, which normally suffers from an upper bend starting at small values of grid bias and becoming more pronounced at $V_g=0$. This objectionable feature may, however, be removed by designing the tube in a certain way, and the characteristic then closely resembles that of a triode while retaining all the advantages of a pentode. The following analysis is thus not to be regarded as restricted to triodes, but merely to tubes whose anode-current—grid-volts characteristic resembles those of triodes. If, however, the actual tubes are not triodes the effect of the anode acting through the screen may be regarded as reflecting the load impedance $z_{n'}$ to the screen plane, where it will appear to have the reduced value $z_{n'}/\mu_2$ where μ_2 is the amplification factor of the screen under the given conditions, while V_p will be equal to (screen voltage$+1/\mu_2\times$anode voltage) and will act not exactly at the screen but at a point between screen and anode given by

$$d_2' = d_2 + \frac{a_2 + d_2}{d_2\mu_2}$$

where, d_2 = distance of screen measured from first grid

a_2 = distance between screen and anode.

Likewise, I_p will now include any current flowing to the screen. The above remarks enable us to proceed with the triode analysis bearing in mind the replacement of symbols necessary to convert to a tetrode or pentode analysis. Writing $\bar{\mu}_0=\mu$; $V_g=E_g+v$; $I_p=i_0+i_1+i_2+i_3+\cdots$; $V_p=E_p-ri_0-z_1i_1-z_2i_2-z_3i_3-\cdots$ equation (8) becomes

$$i_0 + i_1 + i_2 + i_3 = K\left[\frac{E_p - ri_0}{\mu} + E_g + v - \frac{(z_1i_1 + z_2i_2 + z_3i_3)}{\mu}\right]^n.$$

We now abbreviate further by writing the zero-order quantity

$$\frac{E_p - ri_0}{\mu} + E_g = V_l \tag{14}$$

where V_l represents the lumped voltage at the diode equivalent surface corresponding to the grid-bias voltage E_g and to the reduced anode voltage E_p-ri_0. We then have

$$i_0 + i_1 + i_2 + i_3 = KV_l^n\left[1 + \frac{\left(v - \frac{z_1i_1}{\mu}\right) - \left(\frac{z_2i_2}{\mu}\right) - \left(\frac{z_3i_3}{\mu}\right)}{V_l}\right]^n$$

$$= KV_l^n \left[\begin{array}{l} 1 + \dfrac{n(\mu v - z_1 i_1)}{\mu V_l} \\ + \dfrac{n(n-1)}{2!} \dfrac{(\mu v - z_1 i_1)^2}{(\mu V_l)^2} - \dfrac{n z_2 i_2}{\mu V_l} - \dfrac{n z_3 i_3}{\mu V_l} \\ + \dfrac{n(n-1)(n-2)}{3!} \dfrac{(\mu v - z_1 i_1)^3}{(\mu V_l)^3} - \dfrac{n(n-1)}{2!} \dfrac{2(\mu v - z_1 i_1)}{(\mu V_l)^2} \end{array} \right]$$

to the degree of accuracy aimed at. The various orders may then be selected as follows (remembering (10) and the fact that μ variation need not be included explicitly)

$$\left.\begin{aligned} i_0 &= KV_l^n = I_p \\ i_1 &= \frac{\mu v}{r_p + z_1} = \frac{nI_p}{1!}\left(\frac{r_p}{r_p + z_1}\right)\frac{\hat{v}_1}{V_l} \cos \omega t \\ i_2 &= \frac{n(n-1)I_p r_p^3 \cos^2 \omega t}{2!(r_p + z_2)(r_p + z_1)^2}\left(\frac{\hat{v}_1}{V_l}\right)^2 \\ i_3 &= \frac{n(n-1)I_p r_p^4 [(1-2n)z_2 + (n-2)r_p] \cos^3 \omega t}{3!(r_p + z_3)(r_p + z_2)(r_p + z_1)^3}\left(\frac{v_1}{V_l}\right)^3 \end{aligned}\right\} \quad (15)$$

in which we have replaced v by $\hat{v} \cos \omega t$ in order to emphasize the fact that it is according to orders rather than to harmonics that the selection is made. There are, for example, contributions from the second-order to zero-order terms, third-order to first-order, and so forth, so that, to the third-order our actual harmonics are as follows ($\hat{\imath}$ = peak value of i):

$$\left.\begin{aligned} i(h_0) &= I_p + \left[\frac{\hat{\imath}_2}{2}\right]_{z_2 = z_0 = r} \\ i(h_1) &= \left(\hat{\imath}_1 + \frac{3\hat{\imath}_3}{4}\right)_{z_3 = z_1} \cos \omega t \\ i(h_2) &= \frac{\hat{\imath}_2}{2} \cos 2\omega t \\ i(h_3) &= \frac{\hat{\imath}_3}{4} \cos 3\omega t \end{aligned}\right\}. \quad (15a)$$

Particular emphasis is laid upon these considerations in order to assist the understanding of the later parts of the paper.

From (15) and (15a) the ratio of second- and third-harmonic current to fundamental may be obtained for any load impedance, real or complex. It is desirable next to examine the magnitudes of the harmonics in the important case where $z_{n'}$ is substantially independent of n', as, for example, a loud speaker designed to present substantially constant impedance over the range of audio frequencies. Then

$$i(h_2) = \frac{n(n-1)I_p}{4}\left(\frac{r_p}{r_p+z}\right)^3\left(\frac{\hat{v}_1}{V_l}\right)^2 \cos 2\omega t.$$

Assuming $z = r_p$ (maximum output on fundamental), then $i(h_2)$ is given by

$$i(h_2) = \frac{n(n-1)I_p}{32}\left(\frac{\hat{v}_1}{V_l}\right)^2 \cos 2\omega t \tag{16}$$

and the mean power output in watts is, if I_p is in amperes and z in ohms,

$$W(h_2) = \frac{n^2(n-1)^2 I_p{}^2}{2048}\left(\frac{\hat{v}_1}{V_l}\right)^4 z. \tag{17}$$

Now, in the case of $\hat{v}_1 = V_l$, $z = 5000$ ohms, $I_p = 30$ milliamperes, the power outputs are as follows, $n = 3/2$:

$$\left.\begin{aligned} W(h_1) &= 1293 \text{ milliwatts} \\ W(h_2) &= 1.235 \text{ milliwatts} \\ W(h_3) &= 0.0134 \text{ milliwatts} \end{aligned}\right\} \tag{18}$$

from which the following conclusions may be drawn: (a) third-harmonic distortion is *per se* too small to affect quality and (b) that as little as 3.2 per cent total harmonic distortion on voltage output is obtained in the case $z = r_p$, $v_1 = V_l$. But it will now be seen that these criteria are of little value in determining actual distortion in most cases.

V. Distortion by Intermodulation

This phenomenon has until quite recently been much neglected in the design of audio-frequency amplifiers. The writer is indebted to A. C. Bartlett[14] for the following very clear distinction between the well-known phenomenon of cross modulation[15] and that of intermodulation. Bartlett takes cross modulation to be as follows: Suppose a

[14] A. C. Bartlett, "The calculation of modulation products," *Phil. Mag.*, ser. 7, vol. 16, p. 845; October, (1933); part II, *Phil. Mag.*, ser. 7, vol. 17, p. 628; March, (1934). See also D. C. Espley, "Harmonic production and cross modulation in thermionic valves with resistive loads," Proc. I.R.E., vol. 22, pp. 781–790; June, (1934).

[15] S. Ballantine and H. A. Snow, "Reduction of distortion and cross talk in radio receivers by means of variable-mu tetrodes," Proc. I.R.E., vol. 18, pp. 2102–2127; December, (1930).

carrier ω_1 modulated by an audio-frequency Ω, i.e., $(1+k \cos \Omega t) \cos \omega_1 t$, is acting on a tube at the same time as another carrier ω_2, then cross modulation occurs if the output from the tube contains a term of the form $(1+k' \cos \Omega t) \cos \omega_2 t$, i.e., if the audio-frequency modulation of ω_1 is impressed on ω_2. It is easily shown that this will only occur if the characteristic of the tube exhibits third-order curvature. Intermodulation, on the other hand, is taken to denote the production of difference frequencies of any order; a parabolic curve produces only first-order differences, of the general form $nf_1 \pm mf_2$ where n, m are integers; if there are three frequencies present, the differences $pf_2 \pm qf_3$, $rf_1 \pm sf_3$ are obtained in addition, but no second-order differences of the form $nf_1 \pm mf_2 \pm pf_3$. Cross modulation is thus a particular case of the more general phenomenon of intermodulation. If the dynamic characteristic of the tube is parabolic, cross modulation will not arise. The differences produced by intermodulation in audio-frequency amplifiers have been termed *side tones*.[16] These have recently been the subject of an experimental study by J. Owen-Harries.[17] In this case the carrier frequency does not have to be considered, so that m and n are both unity for the first side tones and either 1 or 2 for the second side tones, and so on.

Now Bartlett's classical treatment of the mathematics of modulation products[14] covers all cases where a resistive load is employed, but does not appear to lend itself to a more general treatment where, for example, both tube and load may not be purely resistive. Moreover, the mathematics is probably too difficult for the average engineer. In the sequel a greatly simplified treatment will be given in which only the first and second side tones (arising from second- and third-order characteristic curvature) are calculated for a pair of input frequencies. Bartlett[18] and Harries[17] have pointed out that for the purposes of a simple distortion criterion, two frequencies are sufficient. The reader is referred to these nonmathematical treatments, which render it entirely unnecessary to add anything further in the way of verbal explanations.

In the following analysis we shall content ourselves with the dynamic tube characteristic, as extension to the static tube characteristic may be readily effected by means of the analysis in Section IV. Thus, in the dynamic characteristic,

$$i = a_0 + a_1 v + a_2 v^2 + a_3 v^3 + \cdots. \tag{13a}$$

[16] Little likelihood of confusion with "side tone" in telephony is expected, owing to the difference in context.

[17] J. Owen-Harries, "Amplitude distortion," *Wireless Eng.*, vol. 14, p. 63; February, (1937).

[18] A. C. Bartlett, "Intermodulation in audio frequency amplifiers," *Wireless Eng. and Exp. Wireless*, vol. 12, pp. 70–74; February, (1935).

Let us replace v by the composite signal

$$\left.\begin{array}{c} Ae^{jmt} + Be^{j(nt-\phi)} \\ e^{jmt} \equiv \cos mt, \qquad e^{j(nt-\phi)} \equiv \cos(nt-\phi) \end{array}\right\}. \qquad (19)$$

where,

In the present connection little confusion should arise from the use of exponential notation, but if any difficulty arises, Section VI, dealing with the conventions to be observed, will be found helpful. The position is that the exponential notation saves considerable time and its use may be described as "almost essential" for the treatment given in Sections VII to XII. For this reason no apology is made for its introduction here, which might otherwise have been regarded as premature. There is, however, the important point that, for the simpler modulation products at any rate, the difficulty with multiple angles, so successfully met by Bartlett[18] in another way, may also be met by means of the technique given in Section VI. Furthermore this technique is devised to cover cases where the load and/or tube have a frequency characteristic.

Equations (19) and (13a) give (writing $t_1 = t - \phi/n$)

$$i = a_0 + a_1(Ae^{jmt} + Be^{jnt_1}) + a_2[A^2(e^{jmt})^2 + 2AB(e^{jmt}e^{jnt_1}) + B^2(e^{jnt_1})^2]$$
$$+ a_3[A^3(e^{jmt})^3 + 3A^2B(e^{jmt})^2(e^{jnt_1}) + 3AB^2(e^{jmt})(e^{jnt_1})^2 + B^3(e^{jnt_1})^3].$$

If A is a strong signal and B a weak one, the important side tones are (compare (46) and (48))

$$a_2 2AB(e^{jmt}\cdot e^{jnt_1}) = a_2\cdot AB[\cos\{(m+n)t'-\phi\} + \cos\{(m-n)t'+\phi\}]$$
$$a_3\cdot 3A^2B(e^{jmt})^2 e^{jnt_1}$$
$$= a_3\cdot 3A^2B[\tfrac{1}{2}\cos nt_1' + \tfrac{1}{4}\cos\{(2m+n)t'-\phi\}$$
$$+ \tfrac{1}{4}\cos\{(2m-n)t'+\phi\}] \qquad (20)$$

where t' in this case is equal to t for a purely resistive load and for low frequencies (coefficients a all purely real); $t_1' = t' - \phi/r$.

While the phase angles of the separate side tones may, as Harries[17] points out, be important, it is worth calling attention to the fact that the ratio of the side-tone pairs to the fundamental ($a_1Be^{jnt_1}$) is independent of ϕ at all values of t. Thus, for the first side tones the ratio is $2(a_2/a_1)Ae^{jmt}$, and for the second side tones the ratio is $(3a_3/2a_1)A^2e^{2jmt}$. Although it is difficult to see what may be the physical significance of this ratio, the "pair" comprising two or more discrete frequencies, it is convenient to use it as a criterion of distortion in the same way as previously the ratio of harmonic to fundamental was employed. Now, while it is, of course, an approximation to take the coefficients a_n to be equal to those obtained from the 3/2-power law without anode load, this procedure will here be adopted simply in order to indicate the order of events.

Then,

$$a_1 = \frac{3a_0}{2V_l}; \qquad a_2 = \frac{3a_0}{8V_l^2}; \qquad a_3 = \frac{-a_0}{16V_l^3},$$

so that the ratios of first and second side tones to fundamental become

$$\frac{Ae^{jmt}}{2V_l} \quad \text{and} \quad \frac{-A^2e^{2jmt}}{16V_l^2}. \tag{21}$$

We note that these ratios are independent of B, the amplitude of the weak signal, as would, however, be expected. It should be noted that (21) includes the effect of both the side tones inherent in each of the expressions (20). If we try to take the ratio of the separate side tones $(2)m+n$ and $(2)m-n$ to the signal we do not obtain simple expressions and the phase angles of these ratios will be different for each side tone. If however both $(2)m+n$ and $(2)m-n$ occur in the audible range of frequencies the criterion adopted is probably the best; in other cases it might be too severe.

In order to see the magnitude of distortion that can occur, let us suppose $A = V_l$ (equation (14)). Then the percentage distortion of the weaker signal is from (21) no less than 50 per cent due to the first side tones $m \pm n$ and $6\frac{1}{4}$ per cent due to the second side tones $2m \pm n$. This means, for instance, that if an orchestra is being reproduced the weaker members will come through badly distorted on loud solo passages unless (a) the grid-bias allowance of the last-stage tube is prohibitively large or (b) unless class B amplification is used. Even here, while the first side tones are substantially eliminated with matched tubes, the second side-tone distortion remains at around 6.3 per cent. The need for a design of tube to minimize both first and second side tones is thus only too apparent. It should however, be mentioned that from Harries' experimental results[17] the second side tones in some cases cease to increase rapidly with peak grid volts above about half full drive, an effect which is not explained by the above analysis and may be attributable to some feature of the anode-circuit load. In some cases the second side tone is apparently nearly constant as the grid swing is increased, while in others it appears to increase according to theory. It would take too long to discuss the many aspects of this problem, and the reader is referred to Harries' article. There is little doubt about the fundamentals of the case, which may be summed up as follows. Harmonics per se contribute a total harmonic distortion in the neighborhood of 5 per cent on full drive, while the side tones arising out of harmonic production can produce (in the case of the first side tone) up to about 50 per cent distortion in the plate current of a tube having negligible anode load; for $z = r_p$ it is probable that this figure is around 25 per cent, which figure is attained in Harries' Fig. 6c for the side tones $(m \pm n)$.

VI. Conventions in Regard to the Use of Exponential Notation

In the sequel where transit time of electrons is a complicating factor the tube impedances are complex, and, while the first-order impedance may be evaluated without undue labor in circular functions, the labor involved in the case of the second and third orders is prohibitive unless complex notation is employed. Actually, the use of complex notation does greatly simplify the first-order analysis also, and once certain elementary principles are mastered all difficulties disappear.

While many aspects of complex algebra may be regarded as beyond the scope of the average engineer, it is only with the complex exponential function that we are here concerned, and this is at once the most important, the most fundamental, and the most simple of naturally occurring functions. The circular functions $\sin x$ and $\cos x$ are merely elements of which the exponential e^{ix} is composed. In consequence, it is cumbersome to employ them in complicated problems, but it will probably be some time before the use of complex notation becomes universal. It is to be hoped that this section, which, however, aims at no completeness, will be of assistance in all classes of problems involving complex impedances, as it is by no means restricted to the particular field dealt with in the sequel. Before proceeding it may be as well to point out that, "modulus" and "argument" are synonymous with the more familiar "peak value" (or "amplitude") and "phase." The word "amplitude" is sometimes used instead of "argument" in mathematical works. In description "phase angle" will be used, but in symbols this is "arg z." Mod z will be written $|z|$

The series

$$1 + \frac{z}{1!} + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots + \frac{z^n}{n!} + \cdots \qquad (22)$$

is absolutely convergent for all values of z and therefore defines a one-valued function for all values of z. This function is called the exponential function and will be written e^z whether z is real, imaginary, or complex. The exponential function is possessed of the important property, for any constant λ, real or complex

$$\frac{d}{dz}(e^{\lambda z}) = \lambda e^{z}. \tag{23}$$

Now suppose x to be any real variable and let us choose z, so that

$$z_1 = \cos x + i \sin x \qquad \text{where} \qquad i = \sqrt{-1}. \tag{24}$$

Then,

$$\frac{dz_1}{dx} = -\sin x + i \cos x = i(\cos x + i \sin x) = iz_1.$$

Hence, in virtue of (23),

$$z_1 = Ae^{ix} \tag{25}$$

where A is a constant whose value is found to be unity by comparison of (24) and (25) in the case $x=0$. Thus we have

$$\cos x + i \sin x = e^{ix}.$$

changing x to $-x$ gives

$$\cos x - i \sin x = e^{-ix}$$

whence,

$$\left.\begin{aligned} \cos x &= \frac{1}{2}(e^{ix} + e^{-ix}) \\ \sin x &= \frac{1}{2i}(e^{ix} - e^{-ix}) \end{aligned}\right\}. \tag{26}$$

Now in physical and engineering problems some labor can be saved by replacing $\cos \omega t$ by $\frac{1}{2}(e^{i\omega t}+e^{-i\omega t})$ but it is far more convenient to adopt the convention

$$\cos \omega t \equiv e^{j\omega t}; \qquad -\sin \omega t \equiv je^{j\omega t} \tag{27}$$

which saves covering two sheets of paper instead of one and which also has other unique advantages, provided, and only provided, certain precautions are observed. The use of j rather than i may conveniently remind us that we are now outside the scope of ordinary textbooks on complex algebra in that the $e^{j\omega t}$ of (27) no longer obeys the simple laws of multiplication and division of quantities raised to the power of an index; thus $(e^{j\omega t})^2$ must be left in *this* form until the final stages of the analysis, and must not be written $e^{2j\omega t}$. A further reason for the use of j is in order to insure absence of confusion with the symbol i used for current.

The simplest application of (27) is to linear equations and here all we require to know is the answer to the question: if our final solution is of the form $ze^{j\omega t}=(a+jb)e^{j\omega t}$ how must this be interpreted in circular functions? The answer is immediately provided by (27). Thus

$$(a + jb)e^{j\omega t} = ae^{j\omega t} + b\cdot je^{j\omega t} = a \cos \omega t - b \sin \omega t. \tag{28}$$

The case of nonlinear equations is more difficult, but some simple cases will first be considered. These particular cases will prove valuable guides to general procedure. Let our final solution contain terms of the type $(a+jb)(e^{j\omega t})^2$, and let this term be expressible as a perfect square:

$$\begin{aligned}(a + jb)(e^{j\omega t})^2 &= \left[(a' + jb')e^{j\omega t}\right]^2 \qquad (29)\\ &= (a' \cos \omega t - b' \sin \omega t)^2\\ &= \frac{(a'^2 + b'^2)}{2} - a'b' \sin 2\omega t + \left(\frac{a'^2 - b'^2}{2}\right) \cos 2\omega t.\end{aligned}$$

While by comparison of the two sides of (29) we may determine

$$\begin{aligned}a'^2 - b'^2 &= a\\ 2a'b' &= b\\ a'^2 + b'^2 &= (a^2 + b^2)^{1/2}.\end{aligned}$$

so that, finally, the following interpretation is obtained

$$(a + jb)(e^{j\omega t})^2 = \tfrac{1}{2}\left[(a^2 + b^2)^{1/2} - b \sin 2\omega t + a \cos 2\omega t\right] \tag{30}$$

which may also be written

$$(a + jb)(e^{j\omega t})^2 = \frac{(a^2 + b^2)^{1/2}}{2}\left[1 + \cos\left(2\omega t + \tan^{-1}\frac{b}{a}\right)\right] \tag{31}$$

showing that the second harmonic contains a phase angle in general, or alternatively that second-order solutions exist in a dimension for which time is measured from a new zero, $t=-\frac{1}{2}\omega \tan^{-1}(b/a)$. If we differentiate (30) to time,

$$2j\omega(a + jb)(e^{j\omega t})^2 = -\omega(b \cos 2\omega t + a \sin 2\omega t) \tag{30a}$$

it will be noticed that the constant $(a^2+b^2)^{1/2}$ has disappeared but that the left-hand side may be written

$$-2\omega(b - ja)(e^{j\omega t})^2 \tag{32}$$

which is of the form $(c+jd)(e^{j\omega t})^2$; i.e., of the same form as that with which we started. Thus some restriction is necessary upon the relation (30) and this may be expressed as follows; namely, that a must be expressible in the form $a_1+a_2(\omega)$ where the part a_1 does not depend on ω. In this case we see that (32) possesses a real part dependent entirely on ω (apart from the remote possibility that b contains a part proportional to $1/\omega$) and cannot therefore be regarded as of the form necessary for the relationship (30). Looked at in another way, we must inspect to determine whether our final solution is of the form $(a+jb)(e^{j\omega t})^2$, in which case (30) must be used, or whether it is of the form $(c+jd)e^{j\omega t}j\omega e^{j\omega t}$ in which case $j\omega$ may be regarded as an operator differentiating the expression with respect to time. A little practice makes such inspection possible at a glance.

Let us now inspect the case where we are dealing with the third-order solution. We shall obtain terms of the type $(a+jb)(e^{j\omega t})^3$ and also other terms. Dealing with this one first, let us take the term to have arisen from a cubing process:

$$(a+jb)(e^{j\omega t})^3 = [(a'+jb')e^{j\omega t}]^3 = (a'\cos\omega t - b'\sin\omega t)^3$$

$$= a'^3\cos^3\omega t - 3a'^2b'\cos^2\omega t\sin\omega t + 3a'b'^2\cos\omega t\sin^2\omega t - b'^3\sin^3\omega t$$

$$= (a'^3 - 3a'b'^2)\left(\frac{\cos 3\,\omega t}{4} + \frac{3\cos\omega t}{4}\right)$$

$$+ (3a'^2b' - b'^3)\left(\frac{\sin 3\,\omega t}{-4} + \frac{3\sin\omega t}{4}\right)$$

$$+ 3a'b'^2\cos\omega t - 3a'^2b'\sin\omega t. \tag{33}$$

By means of the relations

$$\left.\begin{aligned} a'^3 - 3a'b'^2 &= a \\ 3a'^2b' - b'^3 &= b \end{aligned}\right\} \tag{34}$$

the answer can be expressed most neatly in the form

$$\left.\begin{aligned} (a+jb)(e^{j\omega t})^3 &= (a^2+b^2)^{1/2}\left[\tfrac{1}{4}\cos 3\omega t' + \tfrac{3}{4}\cos\omega t'\right] \\ \text{where}\quad t' &= t + \frac{1}{3\omega}\tan^{-1}\frac{b}{a} \end{aligned}\right\} \tag{35}$$

also $(a+jb)j(e^{j\omega t})^3 = (a^2+b^2)^{1/2}\left[-\tfrac{1}{4}\sin 3\omega t' - \tfrac{1}{4}\sin\omega t'\right]$ (35a)

from which it will be noticed that, as in the previous examples, there is now a phase displacement, or alternatively, each of the components exists in a new dimension in which time is measured from a new zero, in this case $t=(-1/3\omega)\tan^{-1}(b/a)$. While the analysis leading to the above has been given but sparsely a geometrical interpretation is given with the example to follow. The third-harmonic component may, if desired, be expressed in the form

$$\frac{a}{4}\cos 3\omega t - \frac{b}{4}\sin 3\omega t. \tag{36}$$

For the determination of modulus and argument it is frequently convenient to use (36) in cases where a and b are themselves functions of ω (compare Fig. 4) when the modulus can be obtained as the envelope of the individual components at values of ω where one of these passes through zero. This procedure saves considerable arithmetic. In a similar way the zeros of the curves give information as to the argument which changes by steps of $\pi/2$ between successive zeros. The phase angle of the (third-order) fundamental contribution need not be evaluated but may be taken from curves for the argument of the triple-frequency term, dividing such arguments by 3.

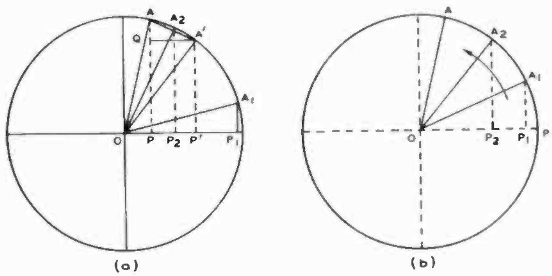


Fig. 2

(a) Construction for determining phase angles $\underline{|A_2OP_2}(=\tan^{-1}b_2/a_2)$ and $\underline{|A_1OP_1}(=\tan^{-1}b_1/a_1)$ from data given in (37) and (38).

(b) Positions of rotating vectors $OA_1(=|a_1+jb_1|e^{j\omega t})$ and $OA_2(=|a_2+jb_2|e^{2j\omega t})$ after rotating from the initial line OP for a time $t=1/2\omega\ \tan^{-1}b_2/a_2 = 1/\omega\ \tan^{-1}b_1/a_1$. In general the vectors would not lie on the same circle, but in all cases we must have $\tan^{-1} P_2A_2/OP_2 = 2\tan^{-1} P_1A_1/OP_1$. Thus (b) is obtained from (a) by moving A_1, A_2 round the circle to the positions shown in (b), when A' will automatically coincide with A_1.

We have next to consider a third-order contribution of the type

$$\begin{aligned}(a+jb)e^{j\omega t}\cdot e^{2j\omega t} &= (a_1+jb_1)e^{j\omega t}(a_2+jb_2)e^{2j\omega t} \qquad \text{(say)}\\ &= (a_1\cos\omega t - b_1\sin\omega t)(a_2\cos 2\omega t - b_2\sin 2\omega t)\\ &= \left.\begin{array}{l}(a_1a_2+b_1b_2)\dfrac{\cos\omega t}{2} - (a_1b_2-a_2b_1)\dfrac{\sin\omega t}{2}\\ +\,(a_1a_2-b_1b_2)\dfrac{\cos 3\omega t}{2} - (a_1b_2+a_2b_1)\dfrac{\sin 3\omega t}{2}\end{array}\right\}. \end{aligned}\tag{37}$$

In this case we have

$$\left.\begin{array}{l} a = (a_1a_2 - b_1b_2)\\ b = (a_2b_1 + a_1b_2)\end{array}\right\} \tag{38}$$

but now there are only two equations and four unknowns. It may, however, be readily shown from the above data that the modulus of fundamental and of triple-frequency terms are equal, while (38) determines the phase angle of the triple-frequency term; i.e., $\tan^{-1}(b/a)$.

In order to find the phase angle of the fundamental component we take advantage of the fact that the modulus is the same for both components. Referring to Fig. 2 (a) let OP represent a, as given above, and let PA be drawn of length b at right angles to OP to meet the circle of radius $(a^2+b^2)^{\frac{1}{2}}$ at the point A. Now from (37) and (38), if ϵ denotes the desired phase angle

$$\tan \epsilon = \frac{a_1 b_2 - a_2 b_1}{a_1 a_2 + b_1 b_2} = \frac{b - 2a_2 b_1}{a + 2b_2 b_1}. \tag{39}$$

We thus proceed along OP and make $PP' = 2b_2 b_1$ and erect the ordinate $P'A'$ to meet the circle in A'. Then since DA' is the modulus of the fundamental component, $\angle A'OP'$ is the phase angle of this component. Hence, making $A'Q$ meet PA at right angles in Q, the point A' may be regarded as having moved around the circle from A, the ordinate being diminished by the amount AQ, i.e., from b to $b - 2a_2 b_1$ (see (39)). We now complete the small triangle AQA' and make OA_2 perpendicular to AA' meeting the circle in A_2. The triangle OP_2A_2 is seen to be similar to the small triangle AQA' so that

$$\frac{P_2A_2}{OP_2} = \frac{QA'}{QA} = \frac{2b_2 b_1}{2a_2 b_1} = \frac{b_2}{a_2}. \tag{40}$$

With the help of the first line of (37) we find

$$\tan^{-1}\frac{b}{a} = \tan^{-1}\frac{b_1}{a_1} + \tan^{-1}\frac{b_2}{a_2} \tag{41}$$

and since $\tan^{-1}(b/a)$ is the angle $\angle AOP_1$, (40) and (41) determine $\tan^{-1}(b_1/a_1)$ as the angle $\angle AOA_2$. We may then construct $\angle A_1OP_1 = \angle AOA_2$, when P_1A_1, OP_1 will represent the values of b_1, a_1, respectively. But, since OA_2 cuts the line AA' at right angles $\angle A_2OA' = \angle A_2OA$, and $\angle AOA' = 2\angle A_2OA = 2\tan^{-1}(b_1/a_1)$. Hence

$$\angle A'OP_1 = \tan^{-1}\frac{b}{a} - 2\tan^{-1}\frac{b_1}{a_1}. \tag{42}$$

We have now obtained the utmost possible information from (37) and (38) but since we require $\angle A'OP_1$ in terms of b/a alone, and since (41) and (42) fail to provide this information, we must seek some restriction on the relative values of a_1, a_2, b_1, and b_2. Such a restriction is readily obtained by reference to Fig 2(b), in which it is seen that whatever the length of the rotors of the rotating vectors $|a_1+jb_1|e^{j\omega t}$, $|a_2+jb_2|e^{2j\omega t}$, the following relationship always holds good:

$$\tan^{-1}\frac{b_2}{a_2} = 2\tan^{-1}\frac{b_1}{a_1}. \tag{43}$$

Eliminating b_2/a_2 between (41) to (43), we obtain

$$\underline{|A'OP_1} = \tan^{-1}\frac{b_1}{a_1} = \frac{1}{3}\tan^{-1}\frac{b}{a}. \tag{44}$$

The points A', A_1 on Fig. 2(a) should then be moved around the circle to meet at the position shown in Fig 2(b), and likewise for the point A_2.

Now, as before, the answer may be expressed in terms of the new time $t'(=t+1/3\omega\tan^{-1}(b/a))$ so that (37) becomes

$$(a+jb)e^{j\omega t}\cdot e^{2j\omega t} = \frac{(a^2+b^2)^{1/2}}{2}(\cos 3\omega t' + \cos\omega t'). \tag{45}$$

also $(a+jb)je^{j\omega t}\cdot e^{2j\omega t} = (a^2+b^2)^{1/2}\left[-\tfrac{1}{2}\sin 3\omega t' - \tfrac{1}{6}\sin\omega t'\right]$ (45a)

Equation (28) together with (30a), (31), (35), (35a), (45), and (45a) provide the necessary basis for the successful handling of problems involving the formation of double- and triple-frequency components by nonlinear apparatus possessing a frequency characteristic. It will be realized that the examples so far cited are simple cases of more general expressions. The six particular cases are nevertheless important guides, in that more general expressions may be formed with their assistance and tested by letting the phase angle tend either to zero or $\pi/2$, and/or writing $\omega_1=\omega_2(=\omega_3)$, or $\omega_2=2\omega_1$, and so forth.

In cases where the frequencies of the constituent vectors are equal, the restriction that these vectors are also cophasal can be shown to be without effect on the higher-frequency component of the resultant, but the lower frequency (or, in the second order, the direct current) component must be multiplied by $\cos\theta$, where θ is the phase angle between the constituent vectors. As, however, the quantity $(a+jb)$ is in general not resolvable into factors at all, other methods are resorted to for finding the component of lower frequency. Such methods are indicated in Appendix III.

It would take too long to extend the above treatment fully to intermodulation and cross modulation, and only some of the more important cases will be mentioned. First, let us consider the simultaneous application of two arbitrarily phased signals to a device in which third-order curvature is unimportant. Intermodulation then produces only first side tones, the required transformation for these being

$$(a + jb)e^{j(\omega_1 t - \phi)}e^{j\omega_2 t}$$
$$= \frac{(a^2 + b^2)^{1/2}}{2}\left[\cos\left\{(\omega_2 + \omega_1)t' - \phi\right\} + \cos\left\{(\omega_2 - \omega_1)t' + \phi\right\}\right] \quad (46)$$

where now

$$t' = t + \frac{1}{(\omega_2 + \omega_1)}\tan^{-1}\frac{b}{a}\cdot \quad (47)$$

We cannot fail to notice the essential similarity of the form of the above expressions to the *third*-order harmonic expressions, yet the above side tones arise only out of *second*-order curvature and are termed the *first* side tones. The form (46) is not, however, dissimilar to the second-order solution (31); the modulus is the same, for the same a and b, and if into the derivation of (31) we write $2\omega = \omega_1 + \omega_2$, then, instead of the constant term we should obtain the difference term in $(\omega_2 - \omega_1)$, and, in short, we should be enabled to arrive at (46) from a knowledge of (31). Then, also, we should be able to arrive at the term giving rise to the second side tone from our knowledge of (35), replacing $(e^{j\omega t})^3$ by $e^{(j\omega_1 t - \phi)}\,(e^{j\omega_2 t})^2$, which gives

$$(a + jb)e^{j(\omega_1 t - \phi)}(e^{j\omega_2 t})^2$$
$$= (a^2 + b^2)^{1/2}\left[\tfrac{1}{4}\cos\left\{(\omega_1 + 2\omega_2)t' - \phi\right\} + \tfrac{3}{4}\cos\left\{(\omega_1 - 2\omega_2)t' - \phi\right\}\right].$$

This answer cannot, however, be correct since we have seen that $(e^{j\omega t})^2$ contains a term independent of t, in which case we expect terms independent of $\omega_2 t$. The correct expression is then found to be

$$(a + jb)e^{j(\omega_1 t - \phi)}(e^{j\omega_2 t})^2 = (a^2 + b^2)^{1/2}\left[\tfrac{1}{2}\cos(\omega_1 t' - \phi)\right.$$
$$\left. + \tfrac{1}{4}\cos\left\{(\omega_1 + 2\omega_2)t' - \phi\right\} + \tfrac{1}{4}\cos\left\{(2\omega_2 - \omega_1)t' + \phi\right\}\right] \quad (48)$$

where

$$t' = t + \frac{1}{\omega_1 + 2\omega_2}\tan^{-1}\frac{b}{a}$$

and it will be observed that the side tones are of equal magnitude, in accordance with experiment.[17] Equation (48) will be found on inspection to reduce to (31) in the case $\omega_1 = 0$; $\omega_2 = \omega$; $\phi = 0$, as well as yielding (35) in the case $\omega_1 = \omega_2 = \omega$; $\phi = 0$. The second side tone is of major importance in commercial pentodes and may exceed the first side tone by many times. The third and higher side tones may also be sufficiently strong to affect reception but as pointed out by Harries a measure of second-side-tone distortion will tell us whether or not this is the case. In the case of cross modulation, if the signal is of the type

$$v = \cos \omega_1 t(1 + k \cos \Omega t) + q \cos \omega_2 t$$

so that, due to third-order curvature of the characteristic we obtain products of powers of the cosines of each of the three frequencies present, it can be shown that the only important third-order terms for obtaining the frequencies $\omega_2 \pm \Omega$ are those in $\cos \omega_1 t \cos \Omega t \cos \omega_2 t$. We therefore have to consider the complex expression.

$$\xi = (a + jb)e^{j\omega_1 t}e^{j\omega_2 t}e^{j\Omega t}$$

which will now be written

$$\xi = [(a_1 + jb_1)e^{j\omega_1 t}e^{j\omega_2 t}](a_2 + jb_2)e^{j\Omega t}$$

which gives with the help of (46) and (28)

$$\xi = \frac{(a_1{}^2 + b_1{}^2)^{1/2}}{2}[\cos(\omega_2 + \omega_1)t_1' + \cos(\omega_2 - \omega_1)t_1'](a_2{}^2 + b_2{}^2)^{1/2} \cos \Omega t_2' \quad (49)$$

where

$$\left\{ \begin{array}{c} t_1' = t + \dfrac{1}{\omega_1 + \omega_2} \tan^{-1} \dfrac{b_1}{a_1}; \quad t_2' = t + \dfrac{1}{\Omega} \tan^{-1} \dfrac{b_2}{a_2} \\ (a_1{}^2 + b_1{}^2)(a_2{}^2 + b_2{}^2) = (a^2 + b^2) \end{array} \right\}.$$

Equation (49) gives rise to the frequency differences $\omega_2 \pm \omega_1 \pm \Omega$ and with the help of first-order terms in ω_1, ω_2 the desired set $\omega_2 - \Omega$, ω_2, $\omega_2 + \Omega$ may be obtained. The only information so far lacking concerns the values of b_1/a_1, b_2/a_2. We know, however, that $\tan^{-1}(b/a) = \tan^{-1}(b_1/a_1) + \tan^{-1}(b_2/a_2)$ and since Ω is an audio frequency it is permissible in most cases to regard b_2 as zero.

Summarizing the remarks of this section, it may be said that the use of the convention

$$e^{nj\omega t} \equiv \cos n\omega t$$

in problems where the input wave is distorted by a nonlinear device possessing in general a frequency characteristic, is perfectly permissible, and in complex problems the labor saving is considerable. Expressions have been developed giving $(e^{nj\omega t})^m(a+jb)$ in real notation, where n is any integer and m has values 1, 2, 3. It should also be mentioned that $\int e^{mj\omega t} \cdot e^{nj\omega t} dt = e^{mj\omega t} \cdot e^{nj\omega t}/(m+n)j\omega$, with reservations when $m = n$. When $m = n$ a term proportional to t would be involved, in cases where we are concerned with expressions of the type $(a+jb)(e^{mj\omega t})^2$, but not in the case of $(a+jb)j(e^{mj\omega t})^2$ which may be written $(a+jb)(je^{2mj\omega t}/2)$.

VII. The Nonlinear Theory of Vacuum Tubes at High Frequencies

Introduction

The sequel extends the electronics of vacuum tubes at high frequencies so as to include the formation of multiple frequencies, arising either solely from finite electron-transit time (temperature-limited case) or from characteristic curvature, including modifications due to electron-transit time (space-charge-limited case). Following the procedure of Sections IV, V, and VI the first, second, and third harmonics will be treated in detail, fourth- and higher-order terms being considered beyond the scope of the present paper. Many applications of the remarks of Section VI will be found, which it is hoped will be of service to those interested in the use of these conventions in other branches of radio.

At the risk of appearing to amplify unduly one phase of a paper not devoted exclusively to transit-time effects, a very brief summary of this field is attempted below. Setting aside the well-remembered work of Hollman[19] and others[20,21] on positive-grid tubes, the papers that have appeared are few in comparison. It is to those investigators of fundamentals as affecting tubes used in circuits with the normal connection that space restricts the sequel.

Early workers who failed to publish their results were B. D. H. Tellegen, R. F. J. Jarvis, and possibly also O. Emersleben. The latter dealt in a remarkable paper[22] with the answer to the question. The n electrodes $E_1, E_2, \cdots, E_n$ are at constant potentials $V_1, V_2, \cdots, V_n$. At the N points $P_1, P_2, \cdots, P_n$ we have the stationary charges $-e_1, -e_2, \cdots, -e_n$. What is the potential $\phi(Q)$ at any given point Q of the space R between the electrodes? In Section II of his paper Emersleben had referred to the "microscopic" aspect. "All the internal actions which are visible by microscopic examination will be sketched in a series of papers." Meanwhile I had made a preliminary announcement of a result of "microscopic" examination in the case of the moving

[19] H. E. Hollman, "The generation and application of the shortest undamped electrical waves," *Hochfrequenztech. und Elektroakustik*, vol. 44, pp. 37–60; August, (1934).

[20] Hannes Alfvèn, *Zeit. für Phys.*, vol. 83, p. 222; June, (1933); Inaugural Dissertation (in English) to the University of Uppsala, Uppsala Universitets Årsskrift, 1934, Matematik och Naturvetinskap. 2., "Investigations on the Ultra-Short Electromagnetic Waves."

[21] R. Cockburn, "Investigations of the fundamental types of electron oscillations in a triode valve," *Proc. Phys. Soc.*, vol. 49, p. 38; January, (1937), and forthcoming paper.

[22] O. Emersleben, "Das Elektrostatischer Feld einer Raumladung-I," *Ann. der Phys.*, vol. 82, p. 713; April, (1927).

charges,[23] and in due course published my analysis in improved form,[24] unaware that I was possibly trespassing on Emersleben's preserves. When in 1928 I came across Emersleben's work it seemed to me that the electrostatic effects, all included automatically[25,26] in my method of approach, need not form a subject of separate investigation. My paper was primarily concerned with the explanation of some experiments I had conducted in 1926 on the frequency characteristic of diode rectification[27] but also gave the first-order solution for the electron velocity in terms of electron-transit time, in closed form (circular functions) and from this all currents, potentials etc., at any part of the diode were readily obtained though not published in full. In 1929 B. D. H. Tellegen pointed out a second-order correction, which I accepted. This authority was the first to convey confirmation of the main features of my 1928 paper (Part II was not published until 1931)[9] in a letter dated December 31, 1931, in which he also gave my results in complex notation. Tellegen also found the result $C=(3/5)C_0$, which I had not mentioned in Part I, though this can be obtained from equations (28) and (29) of Part I.

In September, 1928, J. Müller and F. Tank[28] provided further experiments on diode rectification, using wavelengths down to $\lambda=81$ centimeters.

In August, 1928, a paper appeared from the pen of Stuart Ballantine,[29] in which, *inter alia*, the frequency dependence of the "Schrott ef-

[23] W. E. Benham, "Theory of alternating thermionic currents," paper read before Mathematical and Physical Society, University College, London, January 28, 1927.

[24] W. E. Benham, "Theory of the internal action of thermionic systems at moderately high frequencies—Part I," *Phil. Mag.*, vol. 5, p. 641; March, (1928).

[25] For a formal proof of the equivalence of the "total-current" and "electrical-image" methods, see the first paper of footnote 26.

[26] Papers not dealt with in the survey include the following: W. E. Benham, "Some general relations of vacuum tube electronics," *Wireless Eng. and Exp. Wireless*, 155, vol. 13, pp. 406–413; August, (1936). B. J. Thompson and G. M. Rose, Jr., "Vacuum tubes of small dimensions for use at extremely high frequencies," Proc. I.R.E., vol. 21, pp. 1707–1721; December, (1933). C. Protze, "Raumladungströme bei höhen Frequenzen," *Hochfrequenz. und Elektroakustik*, vol. 42, pp. 20–21; July, (1933). B. Sil, "Variation of interelectrode capacity of a triode at high frequencies," *Phil. Mag.*, vol. 16, pp. 1114–1128; December, (1933). N. E. Lindenblad, "Development of transmitters for frequencies above 300 megacycles," Proc. I.R.E., vol. 23, pp. 1013–1047, September, (1935). C. E. Fay and A. L. Samuel, "Vacuum tubes for generating frequencies above one hundred megacycles," Proc. I.R.E., vol. 23, pp. 199–212; March, (1935). W. A. Krause, "Hochfrequenzmessungen bei 1m. Wellenlange," *Hochfrequenz. und Elektroakustik*, vol. 45, pp. 128–137; April, (1935).

[27] W. E. Benham, "A study of the rectification efficiency of thermionic valves at moderately high frequencies," *Phil. Mag.*, vol. 5, p. 323; February, (1928).

[28] J. Müller and F. Tank, "Über das Verhalten von Glühkathoden gleichrichten bei sehr höhen Frequenzen," *Helvetica Physica Acta*, vol. 1, pp. 447–449; September, (1928).

[29] Stewart Ballantine, "Schrott effect in high-frequency circuits," *Jour. Frank. Inst.*, vol. 206, pp. 159–167; August, (1928).

fect" in a temperature-limited diode was correctly established as a function of the electron-transit time. The technique adopted was unique among transit-time contributions in that use was made of the Rayleigh-Schuster theorem to arrive at the disturbance in the external circuit due to an impressed random disturbance in the charge leaving the cathode per second. Later on attention will be called to the particular term in a general expression for diode impedance which corresponds to the "noise" term established by Ballantine. Although worked out for the temperature-limited diode, Ballantine's result can be shown to hold closely for any degree of space-charge limitation.

During the year 1928, and unknown to myself, R. F. J. Jarvis[30] began improving upon my calculations and presented an important thesis, first in May, 1931, and in final form in October, 1932, on the space currents in two- and three-electrode valves at very high frequencies, with special application to the study of the action of the normal type of oscillator or amplifier at these frequencies. I only came across this authority and his work in 1936, so for present purposes dwell mainly on what I consider to be the most advanced result obtained by Jarvis. In equation (130) of Part II (Part I had dealt mainly with peak voltmeters) Jarvis gives a formula from which the electron damping and capacitance change due to electrons moving between grid and anode of a triode can be obtained to a good approximation. While Jarvis did not calculate the magnitude of the damping effect in this case, he did so in connection with the efficiency of oscillator tubes, obtaining good agreement with the observations of McArthur and Spitzer.[31]

In contrast to my somewhat detailed discussions and to Jarvis' compendious thesis a transit-time paper in miniature was presented by A. Witt, also in 1932, on just two pages of *Comptes Rendus*.[32] Witt establishes negative-resistance regions in the temperature-limited diode, to which I had given insufficient attention and had wrongly based my temperature-limited diode conclusions on the extreme case of a single electron, but the solution given by Witt was only qualitatively correct since the anode current was taken to be equal to the

[30] R. F. J. Jarvis, "The action of thermionic valves and high-frequency oscillators at frequencies above a megacycle—Part II," "A theoretical investigation of the space-charge currents in two- and three-electrode valves at very high frequencies, with special application to the action of the normal type of oscillator or amplifier circuits at these frequencies," Original, May, (1931), revised and extended October, (1932). Approved for the degree of Doctor of Philosophy, 1933. Available in University of London Library.

[31] E. D. McArthur and E. E. Spitzer, "Vacuum tubes as high-frequency oscillators," Proc. I.R.E., vol. 19, pp. 1971–1982; November, (1931).

[32] A. Witt, "Sur l'amorçage des oscillations de très haute frequence," *Comptes Rendus*, vol. 195, pp. 1005–1007, November, (1932).

electron convection current. It is probable that Megaw[33] had a similar solution to that of Witt.

Johannes Müller,[34] unaware of previous work, gave the first correct solution to the temperature-limited case by means of a formula for the impedance of a plane diode which applies everywhere in the temperature-limited region in the case of constant initial velocities (and independent of t), and in the space-charge region in the case of zero initial velocities. In the latter case the solution reduced to that given by me. The almost flawless character of Müller's work, including his demonstration[35] of the oscillation properties of diodes as predicted by theory, marks a fitting close to the first chapter of transit-time theory, which may be described as one in which the fundamentals were established and the basis of multielement-tube theory prepared, though not worked out fully.

The interest in the subject appeared to be on the wane when in November, 1933, the first of a series of papers by F. B. Llewellyn[36] caused one to realize that, in America at any rate, this was not in fact the case. The need for more general boundary conditions with a view to extension to both positive- and negative-grid triodes, was the keynote of Llewellyn's first paper. Parts of the work were of such a nature as to call forth comment from me, and an extensive correspondence followed which led to convergent views. In August, 1934, Llewellyn published an experimental paper[37] relating the phase angle of vacuum-tube transconductance to electron-transit time. Llewellyn's most important achievements on the theoretical side were included in his longest paper, which appeared in October, 1935.[11] The expression given in Appendix I of the present paper reduces to Llewellyn's expression (41) in the case $H=0$. In his most recent (as I write) paper[10] Llewellyn gives a delta network for the plane triode and states, but does not demonstrate, that when a large condenser is placed between cathode and plate the result of computing the input impedance agrees with that formerly presented by him.

[33] E. C. S. Megaw, "Electronic oscillations," *Jour. I.E.E.* (London), vol. 72, p. 313; April, (1933), give good summaries of previous work.

[34] Johannes Müller, "Elektronenschwingungen in Hochvakuum," *Hochfrequenz. und Elektroakustik*, vol. 41, pp. 156–167; May, (1933).

[35] Johannes Müller, "Experimentelle Untersuchungen über Elektronenschwingungen," *Hochfrequenz. und Elektroakustik*, vol. 43, pp. 195–199; June, (1934).

[36] F. B. Llewellyn, "Vacuum tube electronics at ultra-high frequencies," Proc. I.R.E., vol. 21, pp. 1532–1574; November, (1933).

[37] F. B. Llewellyn, "Phase angle of vacuum tube transconductance at very high frequencies," Proc. I.R.E., vol. 22, pp. 947–956; August, (1934).

Meanwhile C. J. Bakker and G. de Vries, in two papers[38,39] published in November, 1934, and July, 1935, had shown that the electrostatic method of approach could, after all, be quite simple. The first of these papers was concerned with the explanation of the ability to obtain amplification by means of a tetrode with a negative plate in which no electrons reached the plate. The method is reminiscent of some work of Sven Benner,[40] who established high-frequency conductance between the deflecting plates of a cathode-ray tube despite no electrons hitting the plates. Incidentally, I must plead guilty to having misunderstood Benner's work in an earlier paper. The reason for this is that as the exact conditions of the problem investigated by Benner were left somewhat to inference (I did not think it fair to tie him down to the arrangement used by L. Bergmann and W. Düring),[41] I concluded that he had in mind the application of his equations also to the case where the electrons were moving normally to the plates. This erroneous conclusion was assisted to some extent by a result I had obtained for this case based on a single electron only, which appeared to give similar results. The discrepancy has been clearly pointed out by D. O. North. In their second paper Bakker and de Vries refine their theory still further and apply it to several important cases of electron damping. Their beautiful result,[42] confirmed experimentally, that this is negative for the cathode-grid space of a negative-grid triode in the temperature-limited condition constitutes the first case of negative conductance due to finite electron-transit time at frequencies low compared with $1/\tau$. The second case appears in Appendix I, already referred to, but here the advantage of a magnetic field makes the phenomenon less striking.

M. J. O. Strutt and A. van der Ziel[43] extended the experiments of Bakker and de Vries and distinguished between electron damping and damping produced by other causes, i.e., cathode-coating resistance and

[38] C. J. Bakker and G. de Vries, "Amplification of small alternating tensions by an inductive action of the electrons in a radio valve with negative anode," *Physica*, vol. 1, pp. 1045–1054. November, (1934).

[39] C. J. Bakker and G. de Vries, "On vacuum tube electronics," *Physica*, vol. 2, pp. 683–697; July, (1935).

[40] Sven Benner, "The change in the dielectric constant of a very rarefied gas by electrons," *Ann. der Phys.*, vol. 3, ser. 7, p. 993; December, (1929).

[41] L. Bergmann and W. Düring, *Ann. der Phys.*, series 5, vol. 1, p. 1041; May, (1929).

[42] See footnote 39, pp. 694 and 695.

[43] M. J. O. Strutt and A. van der Ziel, "Messungen der charakteristischen Eigenschaften von Hochfrequenz-Empfangsröhren zwischen 1, 5 und 60 Megahertz," *E.N.T.*, vol. 12, p. 11; November, (1935). See also "Characteristic constants of h.f. pentodes," *Wireless Eng.*, vol. 14, pp. 478–488; September, (1937), and *Elek. Nach. Tech.*, vol. 14, pp. 281–293; September, (1937).

mutual inductance between tube leads carrying high-frequency currents. W. R. Ferris[44] extends earlier work with B. J. Thompson[45] on electron damping and publishes an account of experiments over a wide range of tubes and conditions. D. O. North[46] supplies a theory to explain these results, the agreement obtained setting a new standard in this field. North used the electrostatic method of approach in dealing with the grid-anode space and the evaluation of the grid-anode displacement current corrected for electron velocities at the grid plane represents an important advance. While the "total-current" method can be made to yield the same results, the "electrostatic" method is of particular value in clarifying and facilitating calculations in multielectrode tubes generally. G. Grünberg in his first paper[47] confirms my results for the plane case and gives a promising beginning to the more difficult problem of cylinders, studied by myself and by Jarvis in the case of an infinitely fine filament. In his second paper[48] Grünberg discusses the sudden application of a surge voltage to a plane diode, and obtains a solution in Bessel functons of the 1/3 order, and purely imaginary argument. In case the surge is a "unit function" the anode current is proportional to t^2 for $0 < t < \tau$.

H. Zuhrt[49] has just published his second paper, in which he points out that by means of North's theory even closer agreement with experiment can be obtained than actually reported by Ferris, due to the fact that for large values of the transit-time ratio (grid-anode transit time÷cathode-grid transit time) extra terms from the exact closed form must be used in the series approximation used for computation. I rather felt that such an attempt to produce closer agreement, though apparently successful, was unjustified without further correction for lead inductance (self- and mutual.)[43]

One or two papers not included in the above account are referred to in the sequel and in the Appendixes, which are entered as such in order to preserve the paper in accordance with the theme "first-, second-, and third-order solution." It is hoped to deal with the literature more fully as opportunity arises.

44 W. R. Ferris, "Input resistance of vacuum tubes as ultra-high-frequency amplifiers," PROC. I.R.E., vol. 24, pp. 82–107; January, (1936).

45 B. J. Thompson and W. R. Ferris, see footnote p. 82 of ref. 44.

46 D. O. North, "Analysis of the effects of space charge on grid impedance," PROC. I.R.E., vol. 24, pp. 108–136; January, (1936).

47 G. Grünberg, "Zur Theorie der Wirkungsweise von Elektronenröhren bei Rasch Veränderlicher Anodenspannung," *Tech. Phys.* (U.S.S.R.), vol. 3, no. 1, pp. 65–80; January, (1936).

48 G. Grünberg, "Über den Anfangsstrom, der durch eine Elektronenröhre beim plötzlichen anlegen einer Stosspannung fliesst.," *Tech. Phys.* (U.S.S.R.), vol. 3, no. 2, pp. 101–110; February, (1936).

49 H. Zuhrt, "Die Leistungsverstörken bei ultrahohen Frequenzen und die Grenze der Rückkapplunggsschuringungen," *Hochfrequenz. und Elektroakustik*, vol. 47, pp. 58 and 79; February, (1936); vol. 49, p. 373; March, (1937).

We now come to consider the technique to be adopted for the problem in hand. Physically speaking, the method comprises both "electrostatic" and "total-current" conceptions. While the mathematical forms favored by Müller and Llewellyn are in many respects an advance on the original Eulerian method which I used, I have never considered the advantage sufficient to warrant any restatement on my part of work already described by the Eulerian technique. The reason for this lies partly in the fact that their method is similar in requiring a knowledge of the variation time before interesting quantities, such as impedance, can be found. In my case it was not the variation time, but the "variation velocity;" but this is not an intrinsic difference. There are some problems for which the Eulerian method is actually more convenient, such, for example as the temperature-limited diode, but space forbids demonstration of this point in the face of a technique which supersedes the above methods both in accuracy and lucidity. The new technique depends on the very simple statement that charge is propagated unchanged, even if the convection current changes. Thus a charge $i_0 dt_0$ arrives as idt in general. While this relation holds for any geometry it will suffice for present purposes to restrict attention to parallel planes. A rigorous proof[50] starting from Benham's equation gives, more explicitly,

$$i_c(x, t) = i_c(0, t_0) \left.\frac{\partial t_0}{\partial t}\right|_x ; \tag{50}$$

i.e., under any conditions of space charge, the electron convection current at any plane x and instant t is related as above to its value at the plane $x=0$ and instant t_0. A similar relation was derived by Jarvis and used for obtaining the result already described as his most advanced, but I had not realized this at the time of making a preliminary announcement.[50] This was partly due to the fact that Jarvis' treatment was in the main similar to the Müller-Llewellyn method; the relation (50) was not employed until late in the thesis (equation (117), p. 71). Likewise, it was not until referring again to Witt's paper that I noticed that his analysis turned on the remark "Si tous les électrons atteignant l'anode y sont captés, $I_a(t)dt = I_0 dt_0$, $I_a(t)$ étant le courant débité sur l'anode." In Witt's case I_0 is the total emission, in Jarvis' case it is the current at the grid surface as determined by diode theory for the cathode-grid space. In both treatments mutual repulsions of electrons were assumed negligible and it is probable that for this reason the most interesting application was missed. To the space-charge-limited case the application of the relation is at once convenient and

[50] W. E. Benham, "Electron inertia as the cause of harmonics in valves," *Nature*, vol. 139, p. 591; April, (1937).

particularly valuable. In the sequel the Jarvis-Witt-Benham relation will be used extensively, as it is seen that the first-order impedance can thereby be obtained without knowledge of the variation time, and the nth harmonic from the $(n-1)$th-order-variation time.

VIII. The Temperature-Limited Diode

When the electric field is unaffected by space charge the equations of motion for a plane geometry in which the x axis is chosen normal to the electrodes, i.e., parallel to the electric field, are

$$\begin{aligned} m\ddot{x} &= e(E_0 + E_1 \cos \omega t) \\ m\ddot{y} &= 0 \\ m\ddot{z} &= 0. \end{aligned} \tag{51}$$

When we adopt our convention $\cos \omega t \equiv e^{j\omega t}$ and integrate once with respect to t over the interval $t-t_0$, we obtain

$$\begin{aligned} \dot{x} &= \dot{x}_0 + \frac{e}{m}\left[E_0(t - t_0) + \frac{E_1}{j\omega}(e^{j\omega t} - e^{j\omega t_0})\right] \\ \dot{y} &= \dot{y}_0 \\ \dot{z} &= \dot{z}_0 \end{aligned}$$

in which $\dot{x}_0 \cdots$ is the electron velocity at the cathode or any arbitrarily chosen origin at the instant t_0, $\dot{x} \cdots$ is the velocity at the value of x reached by electrons leaving the origin $(x=0)$ at the instant t_0 and traveling during a time $t-t_0$. A further integration in similar fashion gives

$$\begin{aligned} x &= \dot{x}_0(t-t_0) + \frac{e}{m}\left[E_0\frac{(t-t_0)^2}{2} + \frac{E_1}{(j\omega)^2}\left\{e^{j\omega t} - e^{j\omega t_0} - (j\omega)(t-t_0)e^{j\omega t_0}\right\}\right] \\ y &= \dot{y}_0(t-t_0) + y_0 \\ z &= \dot{z}_0(t-t_0) + z_0 \end{aligned} \tag{52}$$

giving the position (x, y, z) of a particle which started from the point $(0, y_0, z_0)$ at the instant t_0. Now although in the above development x is seen to be entirely independent of y_0 and z_0, this is merely due to the implicit assumption that the plates are of infinite area. For plates of finite area some particles having tangential emission velocities $\dot{y}_0$, $\dot{z}_0$ escape into the bulb, curve around, and may alight on the side of the anode plate remote from the cathode. Or, if we take x to coincide with the anode, certain of the electrons reaching the plane x do so by an indirect route for which the electric field is everywhere less than the uniform value $E_0+E_1 \cos \omega t$. Oscillations of longer wavelength than three times the wavelength corresponding to a direct transit are reported by C. L. Fortescue (unpublished) which may arise from indirect transits of electrons. The negative resistance ranges predicted for direct transit are the same in space-charge and temperature-limited cases, apart from a difference in electron-transit time. We shall restrict attention in the sequel to the ideal case of infinite planes so that all transits may be considered direct.

To continue with the analysis, if i_0 denotes the total emission, the anode current of our temperature-limited diode is not equal to i_0 at high frequencies. Nor is it equal to $i_0(\partial t_0/\partial t)\big|_x$, but this gives the electron convection current impinging on the anode and may be worked out first, giving from equation (52)

$$i_c = i_0 \frac{\partial t_0}{\partial t}\bigg|_x = i_0 \frac{\dfrac{\dot{x}_0}{\gamma\tau} + 1 + \epsilon e^{j\omega t}\cdot \Upsilon_1(j\omega\tau)}{\dfrac{\dot{x}_0}{\gamma\tau} + 1 + \epsilon e^{j\omega t}\cdot e^{-j\omega\tau}} \tag{53}$$

where

$$\gamma \equiv \frac{e}{m}E_0; \quad \epsilon \equiv \frac{E_1}{E_0}; \quad \tau = t - t_0; \quad \Upsilon_1(j\omega\tau) = \frac{1 - e^{-j\omega\tau}}{j\omega\tau}.$$

Since $(\dot{x}_0 + \gamma\tau)$ represents the final velocity the first term in the numerator and denominator is small compared with unity if the final velocity is large compared with the initial velocity. In other cases we may write $1 + \dot{x}_0/\gamma\tau = 1 + K$ where K will not, however, be independent of t. Then (53) becomes

$$i_c = i_0\left(\frac{1 + \epsilon_1\cdot e^{j\omega t}\Upsilon_1(j\omega\tau)}{1 + \epsilon_1\cdot e^{j\omega t}\cdot e^{-j\omega\tau}}\right) \tag{54}$$

where

$$\epsilon_1 \equiv \frac{\epsilon}{1 + K}.$$

For $0 < |\epsilon_1| < 1$ the term $(1 + \epsilon_1 e^{j\omega t}\cdot e^{-j\omega\tau})^{-1}$ may be expanded by the binomial theorem and the resultant series will be absolutely convergent. Then for $0 < |\epsilon_1| < 1$, we have, writing $\alpha = j\omega\tau$,

$$i_c = i_0[1 + \epsilon_1 e^{j\omega t}\Upsilon_1(\alpha)][1 - \epsilon_1 e^{j\omega t - \alpha} + \epsilon_1^2(e^{j\omega t - \alpha})^2 + \cdots + (-1)^r\epsilon_1^r(e^{j\omega t - \alpha})^r + \cdots]. \tag{55}$$

Now, making use of the identity

$$\left.\begin{aligned} \Upsilon_1(\alpha) - e^{-\alpha} &= \frac{\alpha}{2}\Upsilon_3(\alpha) \\ \text{where,}\qquad \Upsilon_3 &= \frac{2}{\alpha^2}(1 - e^{-\alpha} - \alpha e^{-\alpha}) \end{aligned}\right\} \tag{56}$$

equation (55) may be written

$$i_c = i_0 + i_0\epsilon_1\left(\frac{\alpha}{2}\Upsilon_3\right)e^{j\omega t}\sum_{r=0}^{\infty}\left[-\epsilon_1 e^{(j\omega t-\alpha)}\right]^r. \qquad (55a)$$

Equation (55a) is the exact expression for the wave form of the electron convection current in temperature-limited plane structures for which the degree of temperature limitation is adequate to justify our neglect of space charge in its influence on the electric field, and will not be exact in cases where the anode current is as high as, say, one fourth of the Langmuir value obtainable by raising the available supply of electrons. In order to exemplify the use of the expression let us take ϵ_1 to be independent of time (initial velocities small or negligible compared with final velocity $\dot{x}_0+\gamma\tau$), and let attention be fixed only upon the first three harmonics. For this purpose the expression becomes

$$i_c = i_0\Bigg[1 + \epsilon_1\underrightarrow{\left(\frac{\alpha}{2}\Upsilon_3\right)e^{j\omega t}} - \epsilon_1^2\underrightarrow{\left(\frac{\alpha}{2}\Upsilon_3\right)e^{j\omega t}\cdot e^{j\omega t-\alpha}}$$

$$+ \epsilon_1^3\underrightarrow{\left(\frac{\alpha}{2}\Upsilon_3\right)e^{j\omega t}\cdot(e^{j\omega t-\alpha})^2}\Bigg]. \qquad (57)$$

Now, bearing in mind our remarks in connection with (30a), it will be found to be permissible to take $e^{j\omega t}\cdot e^{j\omega t-\alpha}=e^{-\alpha}(e^{j\omega t})^2$, also $e^{j\omega t}\cdot(e^{(j\omega t-\alpha)})^2=e^{-2\alpha}(e^{j\omega t})^3$. It is necessary to factorize in this manner before proceeding to calculate the harmonic components. The arrows indicate that, since $\alpha=j\omega\times$instantaneous transit time the above terms cannot yet be regarded as harmonics. The term in $e^{j\omega t}$ makes one contribution to the second order and two to the third, while the term in $e^{j\omega t}\cdot e^{j\omega t-\alpha}$ makes one contribution to the third harmonic. The contributions in the order mentioned are

$$\epsilon_1 e^{j\omega t}\alpha_1\frac{d}{d\alpha_0}\left(\frac{\alpha_0}{2}\Upsilon_3\right);\ \left[\frac{\alpha_1^2}{2}\frac{d^2}{d\alpha_0^2}\left(\frac{\alpha_0}{2}\Upsilon_3\right)+\alpha_2\frac{d}{d\alpha_0}\left(\frac{\alpha_0}{2}\Upsilon_3\right)\right]\epsilon_1 e^{j\omega t};$$

$$\alpha_1\frac{d}{d\alpha}\left(\frac{\alpha_0}{2}\Upsilon_3 e^{-\alpha}\right)\epsilon_1^2(e^{j\omega t})^2 \qquad (58)$$

where α_1, α_2 are the first- and second-order components of the variation time $d\alpha/j\omega$ and are given by (neglecting consideration of any direct-current change in transit time)

$$\left.\begin{aligned}\alpha_1 &= -\left(\frac{\alpha_0}{2}\Upsilon_3\right)\epsilon_1 e^{j\omega t}\\ \alpha_2 &= +\left(\frac{\alpha_0}{2}\Upsilon_3\right)\left(e^{-\alpha}-\frac{\Upsilon_3}{4}\right)\epsilon_1^2(e^{j\omega t})^2\end{aligned}\right\}. \qquad (59)$$

So that, in terms of the direct-current transit angle α_0 (see end of bracket) we have

$$i_c = i_0\left[1 + \left(\frac{\alpha}{2}\Upsilon_3\right)\epsilon_1 e^{j\omega t} - \left(\frac{\alpha}{2}\Upsilon_3\right)\left(2e^{-\alpha} - \frac{\Upsilon_3}{2}\right)\epsilon_1^2(e^{j\omega t})^2 + \frac{\alpha}{2}\Upsilon_3\left\{3\left(e^{-2\alpha} - \frac{e^{-\alpha}}{2}\Upsilon_3 + \frac{\Upsilon_3^2}{8}\right) - \frac{3\alpha e^{-\alpha}}{4}\Upsilon_3\right\}\epsilon_1^3(e^{j\omega t})^3\right]_{\alpha=\alpha_0}. \quad (60)$$

Equation (60) may be applied at any point between the plates, and to arrive at the induced current we must take the instantaneous average value of the above over the space between the plates. Thus, in the formula relating the induced current to the electron convection current i_c, namely,[51]

$$i_e = \frac{1}{d}\int_0^d i_c\, dx \qquad (t \text{ constant}). \quad (61)$$

x may be expressed simply in terms of α_0, but with greater difficulty in terms of α, as it would have to be in case we wished to obtain i_e direct from (57). From (52) the direct-current-transit time to the plane x will be defined by

$$x = \dot{x}_0\tau_0 + \frac{eE_0}{2m}\tau_0^2. \quad (62)$$

In the derivation of (59) $\dot{x}_0$ has been taken as zero, but could be retained if desired. It should be noted that τ_0 (the direct-current transit time) is not the same as the value obtained by averaging over a cycle owing to the fact that $\tau_0 \propto v_0^{-1/2}$. There is evidence that this difference gives rise to a rectification effect on the temperature-limited portion of the i_0—v_0 characteristic, but this effect is not considered in the present paper.[52] It may be of interest, however, to mention that experimental evidence of such an effect was obtained by me in 1926 using tubes for which the temperature-limited region of the characteristic was quite flat over the range of v_0 worked with. Later it will be evident that no theory has been published up to the present by which such results could have been explained at an earlier date.

[51] This formula, which was derived by two methods in the first paper of footnote 26, may be shown to be equivalent to that used by Bakker and de Vries, equation (11) of footnote 38 and by North, equation following equation (7) p. 112 of reference 46; namely, $i_e = 1/d\int_0^d i_c(0, t_0)\dot{x}d\tau$, in which τ is the instantaneous electron-transit time. In the problem under investigation, it happens that (61) is the most convenient form of the relation, but where the solution for the induced current is to be applied to interelectrode spaces not comprising a real cathode, the form used by the above-mentioned workers is preferable.

[52] W. E. Benham, "Detection by diodes and triodes at high frequencies," *Wireless Eng.*, vol. 14, pp. 472–477; September, (1937).

In regard to (60) the coefficients of $e^{j\omega t}$ and of $(e^{j\omega t})^2$ lend themselves to integration in closed form, but evaluation of the last term has only been attempted in series form. The general term of such a series would involve triple summation and no attempt has been made to obtain more than three terms in this case. With $dx = eE_0\alpha_0 d\alpha_0/m(j\omega)^2$ we obtain, after integration with respect to x and dividing by the final value of x, writing now α for the final value of $\alpha_0(=j\omega\times$interplate transit time)

$$i_e = i_0\left[1+\left(\frac{\alpha}{3}\Upsilon_4\right)\epsilon_1 e^{j\omega t}-\left(\frac{\alpha}{2}\Upsilon_3^2\right)\epsilon_1^2(e^{j\omega t})^2 + \frac{5\alpha}{16}\left(1-\frac{37\alpha}{15}+\frac{688\alpha^2}{225}-\cdots\right)\epsilon_1^3(e^{j\omega t})^3\right] \quad (63)$$

where

$$\Upsilon_3 = \frac{2}{\alpha^2}(1-e^{-\alpha}-\alpha e^{-\alpha}) = 1-\frac{2}{3}\alpha+\frac{\alpha^2}{4}-\frac{\alpha^3}{15}+\frac{\alpha^4}{72}-\frac{\alpha^5}{420}+\cdots$$
$$= \sum_{m=0}^{\infty}\frac{2(m+1)(-\alpha)^m}{(m+2)!} \quad (64)$$

$$\Upsilon_4 = \frac{6}{\alpha^3}\{\alpha-2+(\alpha+2)e^{-\alpha}\} = 1-\frac{1}{2}\alpha+\frac{3\alpha^2}{20}-\frac{\alpha^3}{30}+\frac{\alpha^4}{168}-\frac{\alpha^5}{1120}+\cdots$$
$$= \sum_{m=0}^{\infty}\frac{6(m+1)(-\alpha)^m}{(m+3)!} \quad (65)$$

$$\Upsilon_3^2 = \frac{4}{\alpha^4}\{1-2(\alpha+1)e^{-\alpha}+(\alpha+1)^2e^{-2\alpha}\}$$
$$= 1-\frac{4}{3}\alpha+\frac{17}{18}\alpha^2+\frac{7}{15}\alpha^3+\frac{43}{240}\alpha^4-\frac{107}{1890}\alpha^5+\cdots$$
$$= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{4(-\alpha)^{m+n}(m+1)(n+1)}{(m+2)!(n+2)!}. \quad (66)$$

It will be noticed that the convergence of Υ_3^2 and also of the third-order series does not appear to be good, but this is deceptive since α is an imaginary quantity. When modulus and argument are taken all series appear to converge quite satisfactorily. Since j occurs throughout (operating on the multiple-angle vectors) the adaptation of (63) requires (28), (30), and (35) to be differentiated with respect to time. When this is done and the result expressed in terms of cosines we obtain, writing $-j\alpha=\xi=\omega\tau_0$,

$$i_e = i_0\left[1 + \frac{\xi}{3}\left|\Upsilon_4\right| \epsilon_1 \cos\left(\omega t + \frac{\pi}{2} + \arg \Upsilon_4\right)\right.$$

$$- \frac{\xi}{4}\left|\Upsilon_3{}^2\right| \epsilon_1{}^2 \cos\left(2\omega t + \frac{\pi}{2} + 2 \arg \Upsilon_3\right)$$

$$+ \frac{5}{32}\xi\left(1 - \frac{7}{450}\xi^2 + \cdots\right)\epsilon_1{}^3\left\{\cos\left(\omega t + \frac{\pi}{2} - \frac{37}{45}\xi + \cdots\right)\right.$$

$$\left.\left. + \cos\left(3\omega t + \frac{\pi}{2} - \frac{37}{15}\xi + \cdots\right)\right\}\right] \tag{67}$$

where,

$$\left.\begin{aligned} \left|\Upsilon_4\right| &= 1 - \frac{\xi^2}{40} + \frac{\xi^4}{2240} - \cdots \qquad \arg \Upsilon_4 = -\frac{1}{2}\xi \\ \left|\Upsilon_3{}^2\right| &= 1 - \frac{\xi^2}{18} + \frac{\xi^4}{720} - \cdots \quad 2 \arg \Upsilon_3 = -\frac{4}{3}\xi\left(1 + \frac{\xi^2}{540} + \cdots\right) \end{aligned}\right\} \tag{68}$$

The values of $|\Upsilon_4|$, $|\Upsilon_3{}^2|$ calculated from (68) are accurate to 1 and 3.5 per cent, respectively, for $0<\xi<\pi$, while arg Υ_4 is exact and 2 arg Υ_3 is accurate to 4 per cent for $0<\xi<6$. Greater accuracy would be unnecessary for most purposes but in special cases extra terms may be required. These may be obtained from (65) and (66).

Let us now apply the above results to the particular case evaluated by Sloane and James by a graphical method. Their values were $\xi=\pi/2$, $v_0=2.82$, $\epsilon_1=1/2$. The value of i_0 need not concern us provided it is well below the space-charge saturation value (44 microamperes in the case $v_0=2.82$ volts). Sloane and James' figure was 0.159 microampere centimeter^{-2}, corresponding to the number of electrons being taken as 40 per cycle. Placing $\xi=\pi/2$; $\epsilon_1=1/2$ we obtain

$$i_e = i_0[1 + \{0.245_5 \cos(\omega t + 45°) + 0.0295 \cos(\omega t + 16°)\}$$
$$- 0.0857 \cos(2\omega t - 30°1') - 0.0295 \cos(3\omega t + 48°)] \tag{69}$$

in which the third-order contribution to the fundamental has been shown separately, the resultant fundamental component being $0.272 \cos(\omega t+42°)$. The moduli of the second and third harmonic may be in error by 1 per cent, but not more in this example, while the phase of the third harmonic is not likely to be more than 5 degrees in error. The phase of the second harmonic would be accurate to a few minutes. Comparing with Sloane and James' figures, namely,

$$i_e = i_0[1 + 0.26 \cos(\omega t + 42°) - 0.092 \cos(2\omega t - 39°)$$
$$- 0.032 \cos(3\omega t + 41°)] \tag{69a}$$

it is seen that the phase angle 42 degrees agrees perfectly; elsewhere agreement is as good as could be expected considering the difficulties of graphical integration and the probable error involved in taking only 40 electrons per cycle. Equation (67) may be used for the estimation of harmonic content at any frequency for values of α up to any desired value, though calculation of the third harmonic for $\alpha > 2j$ requires additional terms for greater accuracy, unless integration in closed form proves to be possible.

IX. Small-Signal Theory of the Space-Charge-Limited Diode

The temperature-limited diode is of little interest for most purposes in comparison with the space-charge-limited diode. In this case it will be seen that all integrals including those in the third-order terms have been evaluated. It is noteworthy that if an integral, however complicated, cannot be evaluated in closed form in the theory of the space-charge-limited diode, then there is a mistake somewhere. This conclusion, combined with previous experience of the magnetron problem, enables one to say with confidence that the transit-time solution of the space-charge-limited plane diode is in exponential functions, not merely, as regards the first order, but for higher orders also, including the case where a magnetic field of constant value is present. This is only one of the many ways in which solutions may be checked. The series expansions are another valuable guide. For example, the coefficients may be plotted independent of sign at equal intervals and a mistake may be detected by the inability to draw a smooth curve between the points. To take a concrete example, in (105) I at first had the coefficient of α^3 as 37/210 and discovered my error by applying this check.

There is no need to restrict ourselves to a single input signal (See also Section XII). By means of the new technique it is particularly easy to include the effect of several signals acting simultaneously. In order, however, to appreciate the essential simplicity of the technique, and also to point the differences in the procedure in the case when space charge is not negligible, let us consider first a total current of the form $\bar{\imath}_0 + \bar{\imath}_1 \cos \omega t$ and seek the value of the applied potential necessary to give rise to this current. With the following equation[42] as a starting point currents are in amperes per square centimeter of plane anode, and potentials (when obtained) in volts:

$$\dddot{x} = 2.00_2 \times 10^{28}(\bar{\imath}_0 + \bar{\imath}_1 e^{j\omega t}) \quad (\text{cm sec}^{-3}) \tag{70}$$

in which the numeric represents the value of $4\pi ec^2/10m$ in erg centimeters coulomb^{-1} gramme^{-1}. We now integrate (70) with respect to t, and obtain the acceleration of the electrons at any plane x, at an instant $(t-t_0)$ seconds later than the instant of departure of the electrons from the cathode surface. In other words, the integration extends only over the interval $(t-t_0)$ and gives, subject to $\ddot{x}=\ddot{x}_0$ at the cathode,

$$\ddot{x} = 2.00_2 \times 10^{28}\left[\bar{\imath}_0(t - t_0) + \frac{\bar{\imath}_1}{j\omega}(e^{j\omega t} - e^{j\omega t_0})\right] + \ddot{x}_0. \quad (70a)$$

The right-hand side of (70) can be shown to be proportional to the electric field, but it is slightly inconvenient to work with the electric field since this is, unlike $\bar{\imath}_0$, not independent of t. Equation (70a) is, of course, the same as (51) if $\boldsymbol{E}_0$ and $\boldsymbol{E}_1$ are now regarded as functions of $t-t_0$. The velocity acquired by the electrons leaving the cathode at the instant t_0 with the emission velocity $\dot{x}_0$ is obtained by a further integration in similar fashion

$$\bar{\dot{x}} = 2.00_2 \times 10^{28}\left[\frac{\bar{\imath}_0}{2}(t - t_0)^2 + \frac{\bar{\imath}_1}{(j\omega)^2}\left\{e^{j\omega t} - e^{j\omega t_0} - j\omega(t - t_0)e^{j\omega t_0}\right\}\right]$$
$$+ \ddot{x}_0(t - t_0) + \bar{\dot{x}}_0 \cdots. \quad (70b)$$

At a given value of t the electrons which left the cathode with a distribution of velocities about $\bar{\dot{x}}_0$ as mean value have velocities distributed about $\bar{\dot{x}}$ as mean value. There is no other approximation involved, provided the transit time between cathode and potential minimum is so short[53] that the potential minimum may be regarded as our source of electrons, and the alternating field is taken as sensibly zero between cathode and potential minimum. This attitude[54] has the advantage that $\ddot{x}_0$ may then be taken as zero, while $\bar{\dot{x}}_0{}^2$ is the same as at the cathode surface. This last conclusion is at first sight somewhat surprising since electrons are in a retarding direct-current field between cathode and potential minimum. It arises, however, from the consideration that the velocity distribution is Maxwellian; the more slowly moving electrons in the emission do not become part of the anode cur-

[53] Note added in proof. This transit time is in many cases of importance. R. Cockburn in a paper entitled "The variation of voltage-distribution and of electron transit time in the space-charge-limited planar diode," *Proc. Phys. Soc.*, vol. 50, pp. 298–310; March, (1938), points out in a footnote to p. 309 that North's upper limit for τ_m/τ (see p. 136 of footnote 46) should be 0.4 rather than 0.04, the error arising from the omission of a factor 10 from equation (313) of a paper by I. Langmuir and K. T. Compton, *Rev. Mod. Phys.*, part II, vol. 3, p. 244; April, (1931). Under typical conditions τ_m/τ_{k_0} for a negative-grid triode is 20 per cent, and it seems that the next step in transit-time theory lies in the corrections which must be made in order to allow for conditions between the cathode and potential minimum. Cockburn's work forms a satisfactory direct-current basis for such a development.

[54] There are many fine points in connection with the choice between cathode and potential minimum as source, but space forbids detailed examination here.

rent but return to the cathode, and the average kinetic energy of the faster electrons is raised on this account to an extent which exactly compensates for the loss of energy sustained in virtue of the retarding field. We may define $\bar{\dot{x}}_0$ as the arithmetic mean velocity which may be arrived at, if required, from the relation $\bar{\dot{x}}_0 = \sqrt{\pi kT/2m}$, where T is the cathode temperature in degrees Kelvin and $k = 1.372 \times 10^{-16}$ ergs per degree. These conclusions, however, do not apply on the Fermi-Dirac statistics, but are sufficiently accurate in most cases (space-charge density is low compared with conditions inside metal).

A further integration over the interval $t - t_0$ gives the mean position $\bar{x}$ of electrons starting from the cathode with the mean velocity $\bar{\dot{x}}_0$.

$$\bar{x} = 2.00_2 \times 10^{28}\left[\frac{\bar{\imath}_0}{6}(t - t_0)^3 + \frac{\bar{\imath}_1}{(j\omega)^3}\left\{e^{j\omega t} - e^{j\omega t_0} - j\omega(t - t_0) - \frac{(j\omega)^2}{2}(t - t_0)^2\right\}\right] + \frac{\ddot{x}_0}{2}(t - t_0)^2 + \bar{\dot{x}}_0(t - t_0). \quad (70c)$$

The mean here is to be distinguished from the mean used later on to denote time average. If $\bar{\imath}_0$ is unchanged by the presence of the signal, we also have the direct-current relation

$$\bar{x} = 2.00_2 \times 10^{28}\frac{\bar{\imath}_0}{6}\tau_0^3 + \frac{\ddot{x}_0}{2}\tau_0^2 + \bar{\dot{x}}_0\tau_0 \quad (70d)$$

where τ_0 is the value assumed by $(t - t_0)$ in the absence of signal. The difference between (70c) and (70d) may be taken and the change in $(t - t_0)$ calculated. We thus arrive at the various components of the variation time $(\tau - \tau_0)$. These may then be used to obtain $\ddot{x}$ (equation (70a)) in a form suitable for integration with respect to x, and the potential obtained by means of such integration. While this method has been favored by a number of workers, including Jarvis,[30] Müller,[34] and Llewellyn,[11] the physics of the problem is brought out better by the $\partial t_0/\partial t$ method, which also simplifies the calculations somewhat. The electron convection current is then obtained in terms of the instantaneous transit time in the first instance, and if we are only concerned with small signals there is no necessity to calculate the variation time at all. From (70c), in which $\ddot{x}_0$, $\bar{\dot{x}}_0$ may in general be functions of t_0, we readily obtain (omitting the numerical constant)

$$\left.\frac{\partial t_0}{\partial t}\right|_x = \frac{\bar{\dot{x}}_0 + \tau\ddot{x}_0 + \dfrac{\bar{\imath}_0\tau^2}{2} + \dfrac{\bar{\imath}_1}{(j\omega)^2}\left\{e^{j\omega t} - e^{j\omega t_0} - j\omega\tau e^{j\omega t_0}\right\}}{\bar{\dot{x}}_0 + \tau\left(\ddot{x}_0 - \dfrac{\partial\bar{\dot{x}}_0}{\partial t_0}\right) + \dfrac{\tau^2}{2}\left(\bar{\imath}_0 + \bar{\imath}_1 e^{j\omega t_0} - \dfrac{\partial\ddot{x}_0}{\partial t_0}\right)} \quad (71)$$

where $\tau = t - t_0$ and $\partial \ddot{x}_0/\partial t$ has been taken as $\partial \ddot{x}_0/\partial t_0 \cdot \partial t_0/\partial t$ in order to bring it into the position shown, and, moreover, for the additional reason that $\ddot{x}_0$, $\bar{\dot{x}}_0$ are functions of t_0. Now $(\bar{\imath}_0 + \bar{\imath}_1 e^{j\omega t_0}) - \partial \ddot{x}_0/\partial t_0$ represents the difference between the total current at the potential minimum and the displacement current existing there; i.e., the electron convection current at the plane $x=0$ at the instant t_0. For values of $\bar{\dot{x}}_0 + \ddot{x}_0\tau$ up to 10 per cent of $i_0(\tau^2/2)$ the error in dropping these terms out of both numerator and denominator is for small signals only about 1 per cent in $\partial t_0/\partial t$. Moreover $\partial \bar{\dot{x}}_0/\partial t_0$ can be dropped as a good approximation in the case of either an actual cathode, or a potential minimum in front of same (since, neglecting alternating-current fields between cathode and potential minimum, $\dot{x}_0$ is the same, as we have seen, at both places), *provided* we are not concerning ourselves with thermal fluctuations but are investigating only the effect of fluctuations imposed by the signal applied between the electrodes. Thus we have (with an error not greater than 2 per cent for interelectrode potentials of 15 volts or over)

$$\left.\frac{\partial t_0}{\partial t}\right|_x = \frac{\bar{\imath}_0 + \dfrac{2\bar{\imath}_1 e^{j\omega t}}{(j\omega\tau)^2}\left\{1 - e^{-\alpha} - \alpha e^{-\alpha}\right\}}{\left\{\bar{\imath}_0 + \bar{\imath}_1 e^{j\omega t_0} - \dfrac{\partial \ddot{x}_0}{\partial t_0}\right\}} = \frac{\bar{\imath}_0 + \bar{\imath}_1 \Upsilon_3(\alpha) e^{j\omega t}}{i_c(0, t_0)}. \tag{72}$$

We now see that the electron convection current at the plane x, instant t is given by the numerator of the expression, so that

$$i_c = \bar{\imath}_0 + \bar{\imath}_1 e^{j\omega t}\Upsilon_3(\alpha). \tag{73}$$

We see from (73) that the electron convection current has been obtained in terms of the instantaneous transit time. In order to obtain this current in terms of the direct-current transit time it is no longer necessary to make use of the "variation time" unless we are dealing with higher orders than the first. Thus for small signals all that is necessary is to write $\alpha = \alpha_0$

$$i_c = \bar{\imath}_0 + \bar{\imath}_1 e^{j\omega t}\Upsilon_3(\alpha_0). \tag{73a}$$

The method by which I originally arrived at the result (73a) involved first finding the electron-velocity ripple, or "variation velocity" in terms of the direct-current transit time, and subsequent workers have arrived at the solution by means of the "variation time." It is, however, only comparatively recently that I have arrived at (73), and from the physical point of view it is very satisfactory that the solution does, after all, come out in this way. For, when further components are included, the analysis leading to (73) is in no way complicated, we obtain, instead of (73),

$$i_c = \bar{\imath}_0 + \sum_{n=1}^{\infty} \bar{\imath}_n e^{nj\omega t} \Upsilon_3(n\alpha)$$
$$= \bar{\imath}_0 + \bar{\imath}_1 \Upsilon_3(\alpha) e^{j\omega t} + \bar{\imath}_2 \Upsilon_3(2\alpha) e^{2j\omega t} + \bar{\imath}_3 \Upsilon_3(3\alpha) e^{3j\omega t} + \cdots . \qquad (74)$$

Equation (74) forms a most convenient starting point for further discussions. Thus, in terms of the direct-current transit time (74) becomes, for *small* signals (i.e., the terms are *not* harmonics produced by the tube)

$$i_c = \bar{\imath}_0 + \bar{\imath}_1 \Upsilon_3(\alpha_0) e^{j\omega t} + \bar{\imath}_2 \Upsilon_3(2\alpha_0) e^{2j\omega t} + \bar{\imath}_3 \Upsilon_3(3\alpha_0) e^{3j\omega t} + \cdots . \qquad (74a)$$

The current induced in the electrodes is then (still neglecting initial velocities)

$$i_e = \bar{\imath}_0 + \bar{\imath}_1 \Upsilon_4(\alpha_0) e^{j\omega t} + \bar{\imath}_2 \Upsilon_4(2\alpha_0) e^{2j\omega t} + \bar{\imath}_3 \Upsilon_4(3\alpha_0) e^{3j\omega t} + \cdots . \qquad (75)$$

The potential giving rise to these conditions may now be found by taking advantage of the fact that the cold-capacitance current is equal to the difference between total current and induced current. The assumed total current is given by

$$\bar{\imath}_0 + \bar{\imath}_1 e^{j\omega t} + \bar{\imath}_2 e^{2j\omega t} + \bar{\imath}_3 e^{3j\omega t} + \cdots . \qquad (75a)$$

We thus obtain (C_0 = cold capacitance)

$$C_0 \frac{dv}{dt} = \bar{\imath}_1 [1 - \Upsilon_4(\alpha)] e^{j\omega t} + \bar{\imath}_2 [1 - \Upsilon_4(2\alpha)] e^{2j\omega t}$$
$$+ \bar{\imath}_3 [1 - \Upsilon_4(3\alpha)] e^{3j\omega t} + \cdots \qquad (76)$$

or, since,

$$1 - \Upsilon_4(n\alpha) = \frac{n\alpha}{2} \Upsilon_6(n\alpha) \quad \text{(see (106))},$$

$$C_0 \frac{dv}{dt} = \frac{\alpha}{2} \Upsilon_6(\alpha) \bar{\imath}_1 e^{j\omega t} + \alpha \Upsilon_6(2\alpha) \bar{\imath}_2 e^{2j\omega t} + \frac{3\alpha}{2} \Upsilon_6(3\alpha) \bar{\imath}_3 e^{3j\omega t} + \cdots \qquad (77)$$

giving, after dividing through by $\tau_0/2$ and integrating with respect to t

$$\frac{2C_0 v}{\tau_0} = \bar{\imath}_0 + \bar{\imath}_1 \Upsilon_6(\alpha) e^{j\omega t} + \bar{\imath}_2 \Upsilon_6(2\alpha) e^{2j\omega t} + \bar{\imath}_3 \Upsilon_6(3\alpha) e^{3j\omega t}. \qquad (78)$$

where $\bar{\imath}_0$ is the constant of integration. Similarly, if the input signals should happen to be of frequencies unconnected by any integral relation, we obtain in similar fashion

$$\frac{2C_0 v}{\tau_0} = \bar{\imath}_0 + \sum_1^{\infty} \bar{\imath}_n \Upsilon_6(\alpha_n) e^{j\omega_n t}. \qquad (79)$$

The above illustrates the simplest possible application, and only applies first, for initial velocities small and second, signals small compared with the potential between the plates, and for space-charge-limited conditions.

Before proceeding to the case of large signals attention is again called to the question of allowing for initial velocities; we have already seen the extent to which i_c is free from error. In arriving at (75) for i_c it was tacitly assumed that $\ddot{x}_0$, $\ddot{x}$ were negligible in (70d). Nevertheless the error in omitting these terms during the integration of i_c with respect to x is balanced, since the integration is merely an averaging process and involves dividing by x after the integration is complete $(x=d)$, so that the error involved is estimated at *not more than* 2 per cent in the case of i_s for interelectrode potentials of 15 volts.

The error in (78) is, however, in excess of that for i_e since we have taken the difference between two nearly equal quantities, one of which is exact $(\bar{\imath})$ and the other of which is in error (i_e). We have, at the same time, not used the relation

$$g_0\tau_0 = 2C_0 \tag{80}$$

which may be in error by some 50 per cent due to initial velocities. Let us, therefore not use this relation, which depends on the 3/2-power law neglecting initial velocities. The main source of error lies probably in that a close estimate of τ_0 is not a simple matter, even with the information provided by the work of R. Cockburn.[21,53] For, in application to triodes, the transit times must be regarded as having a distribution, not only due to initial velocities but also due to the corrugations of the grid. The analysis given in Section I suggests the necessary correction factor as $((\gamma+\mu)/\mu)^{2/3}$ as applied to τ_0 and $(\mu/(\gamma+\mu))^{4/3}$ as applied to v, the currents being supposed the same. The above correction factors involve that the signal applied to the grid will have to be $((\gamma+\mu)/\mu)^{2/3}(v-v_0)$, τ_0 then requiring no correction if calculation is based upon a diode coinciding with cathode grid. If, however, τ_0 is the *actual* transit time to the grid plane of the triode then the correction factor $t_0(v-v_0)$ is $((\gamma+\mu)/\mu)^{4/3}$, which is quite an important correction and substantially equivalent to North's[46] figure for this case.

In cases where the inclusion of initial velocities is desirable (71) can be made to yield the alternating-current equation (simple small input signal)

$$\bar{\imath}_c(x, t) = \bar{\imath}_c(0, t_0)\left[\frac{\dot{x}_0 + \ddot{x}_0\tau_0}{\dot{x}}\right] + \left[\frac{1.00_1 \times 10^{28}}{\dot{x}}\bar{\imath}_0\tau_0{}^2\right]\bar{\imath}_1\Upsilon_3(\alpha) + \left[\frac{i_0\tau_0}{\dot{x}}\right]\frac{\partial\dot{x}_0}{\partial t} \tag{82}$$

where all quantities in square brackets are direct-current terms.[55] Equation (82) is in a form suitable for integration with respect to x, since in all cases $\dot{x}$ appears in the denominator and $dx = [\dot{x}d\tau]$. Since the integration must take place at t constant we must first write

$$\left.\begin{aligned} \bar{\imath}_c(0, t_0) &= \bar{\imath}_c(0, t)\cdot e^{-\alpha} \\ \frac{\partial \dot{x}_0}{\partial t} &= \frac{\partial}{\partial t}\dot{x}(0, t)\cdot e^{-\alpha} \end{aligned}\right\} \qquad (83)$$

where $(0, t)$ signifies that these quantities are reckoned at the cathode, not at the instant the electrons constituting $\bar{\imath}_c(x, t)$ leave the cathode, but at the instant when these reach the plane x; $\bar{\imath}_c(0, t)$ thus corresponds to a later set of electrons than $\bar{\imath}_c(x, t)$. The integration involves the following integrals:

$$\left.\begin{aligned} \int_0^{\alpha_0} e^{-\alpha}d\alpha &= 1 - e^{-\alpha_0} = \alpha_0\Upsilon_1(\alpha_0) \\ \int_0^{\alpha_0} \alpha e^{-\alpha}d\alpha &= \frac{\alpha_0^2}{2}\Upsilon_3(\alpha_0) \\ \int_0^{\alpha_0} \alpha^2\Upsilon_3(\alpha)d\alpha &= \frac{\alpha_0^3}{3}\Upsilon_4(\alpha_0) \end{aligned}\right\} \qquad (84)$$

All of which are of a simple nature, characteristic of most integrations met with in space-charge-limited tubes with plane geometry. The expression for i_e is thus readily arrived at and from this expression the potential may be obtained

$$v = \frac{\bar{\imath}_1}{j\omega C_0}\left[1 - \frac{2.00_2\times 10^{28}\bar{\imath}_0\tau_0^3}{6d}\Upsilon_4(\alpha_0) - \frac{\bar{\imath}_c(0, t)}{\bar{\imath}_1 d}\left\{\frac{\ddot{x}_0\tau_0^2}{2}\Upsilon_3(\alpha_0) + \frac{\dot{x}_0\tau_0}{1}\Upsilon_1(\alpha_0)\right\} - \frac{\bar{\imath}_0\tau_0^2}{2\bar{\imath}_1 d}\Upsilon_3(\alpha_0)\frac{\partial}{\partial t}\dot{x}(0, t)\right]. \qquad (85)$$

[55] Note added in proof. It is important to note that with (71), (72), and (82) we are now free from my 1928–1931 assumption that the electron-convection-current ripple at the cathode is equal to the total current ripple. In fact, $i_c(0, t_0)$ may be entirely direct current and $\bar{\imath}\Upsilon_3$ will still be the important term in the solution for $\bar{\imath}_c(x, t)$. If noise fluctuations are included, the first and last terms of (82) predominate at points close to the cathode (τ_0 small). This must not, however, be taken as a point against close-clearance structures, since for these $\bar{\imath}_1$ will be correspondingly large. The inability of electrons to respond to impressed forces immediately on emerging from a real cathode (see p. 8 of reference 60) is analogous to the inertial manifestations of electrons in high-vacuum cathode-ray tubes, in which electrons obey optical laws, and are prevented by their inertia from following exactly the lines of force in the varying fields on an electron-lens system.

Equation (85) is equivalent to Llewellyn's[11] (41) and is in a form which permits analysis of the physical significance of each term. Thus $1/j\omega C_0$ gives the cold-capacitative impedance of the tube, the second term is the principal term affecting the impedance when space current $\bar{\imath}_0$ is taken, the subsequent terms indicate a lowering of the impedance due to electron convection current entering at $x=0$ (or a rise if for any reason $i_c(0, t)$ is in antiphase with $\bar{\imath}_1$) and due to fluctuations[56] in the velocity $\dot{x}(0, t)$. In Appendix I will be found a still more general expression than (85) applicable in case a magnetic field tilted at any angle is applied, and it is shown that the magnetron diode can behave as a negative resistance in series with a positive inductance at zero frequency. In Section X we come to consider large amplitudes. Initial velocities will be neglected in order to simplify the analysis, not because they are less important than in the case of small signals.

X. Large-Signal Theory of a Space-Charge-Limited Diode Neglecting Initial Velocities

We use (74) as our starting point, where now the different components must be regarded as due to characteristic curvature as well as to the presence of harmonics in the input potential. For the time being we shall restrict our attention to an input signal all of whose components are integral multiples of $\omega/2\pi$, such as might correspond, for example, to a single carrier and its harmonics produced by distortion of the wave form arising from the circulation of harmonic currents through the input circuit components.

Rewriting the equation, for convenience,

$$i_c = \bar{\imath}_0 + \underrightarrow{\bar{\imath}_1 \Upsilon_3(\alpha) e^{j\omega t}} + \underrightarrow{\bar{\imath}_2 \Upsilon_3(2\alpha) e^{2j\omega t}} + \underrightarrow{\bar{\imath}_3 \Upsilon_3(3\alpha) e^{3j\omega t}} + \cdots. \tag{74b}$$

The arrows possess a similar significance to that already explained for the temperature-limited diode. Variable transit time gives rise to higher-order terms.

It is most important in the large-signal theory to include in the second-order variation time a term apparently omitted by Llewellyn[11]

[56] It is the last term of (85) which corresponds to Ballantine's "noise" term.[29] Note the presence of the same function, $\Upsilon_3(\alpha)$, as devised by Ballantine for the frequency factor applicable to the Schrott effect in temperature-limited structures. In view of previous remarks in Section IX the frequency-spectrum characteristic of noise, may be represented by simply placing a summation sign in front of this term (compare equation (79)). The remaining terms of (85) can be arranged to refer partly to signal and partly to noise terms. In the space-charge-limited condition, corrections for the potential minimum will tend to invalidate (85), particularly as regards the last two terms. It is thought, however, that the predominant transit-angle function for emission fluctuations will remain as Υ_3, whatever happens to the absolute value of the noise.

(equation (17)), arising from the rectified current δi_0. We then have, for $\dot{x}_0$, $\ddot{x}_0$ zero,

$$\alpha_1 = j\omega\tau_1 = -\frac{\bar{i}_1}{i_0}(\Upsilon_3 - e^{-\alpha_0})e^{j\omega t} \tag{86a}$$

$$\alpha_2 = j\omega\tau_2 = -\left[\frac{\alpha_0\delta i_0}{3i_0} + \frac{\alpha_1 e^{-\alpha_0}\bar{i}_1 e^{j\omega t}}{i_0} + \frac{\alpha_1^2}{\alpha_0}\right] - \frac{\bar{i}_2 e^{2j\omega t}}{2i_0}\left[\Upsilon_3(2\alpha) - e^{-2\alpha}\right] \tag{86b}$$

where $\delta i_0 = \bar{i}_0 - i_0$, i_0 = direct current in the absence of a signal. The last term of (86b) corresponds to the top line of Llewellyn's expression for δ_2, the last term in the first bracket to his term in δ_1^2, the middle term to his term in δ_1. Llewellyn's remaining terms have to do solely with initial velocities and accelerations, and the term in δi_0 was possibly overlooked. This term arises when taking the difference between (70c), in Section IX, and (70d) after allowing for $\bar{i} = i_0 + \delta i_0$ in the former, and replacing $\bar{i}$ by i_0 in the latter equation. Unless this term is included we find that the third-order contribution to the fundamental is nine times the third harmonic instead of only three times (at low frequencies).

It is as well to mention that Υ is used instead of $\Upsilon(\alpha)$ for purposes of abbreviation, but where multiple angles (2α, 3α) occur these are always written out in full. In terms of α_0 (74b) becomes, writing now $\alpha_0 = \alpha$,

$$i_c = \bar{i}_0 + \bar{i}_1\Upsilon_3 e^{j\omega t} + \left\{\alpha_1 \frac{\partial\Upsilon_3}{\partial\alpha}\bar{i}_1 e^{j\omega t} + \bar{i}_2\Upsilon_3(2\alpha)e^{2j\omega t}\right\}$$
$$+ \left\{\bar{i}_1 e^{j\omega t}\left(\frac{\alpha_1^2}{2!}\frac{\partial^2\Upsilon_2}{\partial\alpha^2} + \alpha_2\frac{\partial\Upsilon_3}{\partial\alpha}\right) + \bar{i}_2 e^{2j\omega t}\alpha_1\frac{\partial\Upsilon_3(2\alpha)}{\partial\alpha} + \bar{i}_3 e^{3j\omega t}\Upsilon_3(3\alpha)\right\} \tag{87}$$

in which the values of $\partial\Upsilon_3/\partial\alpha$, $\partial\Upsilon_3(2\alpha)/\partial\alpha$, $\partial^2\Upsilon_3/\partial\alpha^2$ are as follows:

$$\left.\begin{aligned}\frac{\partial\Upsilon_3}{\partial\alpha} &= -\frac{2}{\alpha}(\Upsilon_3 - e^{-\alpha})\\ \frac{\partial\Upsilon_3(2\alpha)}{\partial\alpha} &= -\frac{4}{2\alpha}\left[\Upsilon_3(2\alpha) - e^{-2\alpha}\right]\\ \frac{\partial^2\Upsilon_3}{\partial\alpha^2} &= -\frac{2}{\alpha}e^{-\alpha} + \frac{6}{\alpha^2}(\Upsilon_3 - e^{-\alpha})\end{aligned}\right\}. \tag{88}$$

With the help of (86a), (86b), and (88), (87) becomes

$$i_c = \bar{\imath}_0 + \bar{\imath}_1 e^{j\omega t}\Upsilon_3 + \left\{ \bar{\imath}_2\Upsilon_3(2\alpha)e^{2j\omega t} + \frac{2\bar{\imath}_1^2}{i_0\alpha}(\Upsilon_3 - e^{-\alpha})^2(e^{j\omega t})^2 \right\}$$
$$+ \left\{ \bar{\imath}_3\Upsilon_3(3\alpha)e^{3j\omega t} + \left[\frac{3\bar{\imath}_2 e^{2j\omega t}}{i_0\alpha}(\Upsilon_3(2\alpha) - e^{-2\alpha}) + \frac{2\delta i_0}{3i_0}\right](\Upsilon_3 - e^{-\alpha})\bar{\imath}_1 e^{j\omega t} \right.$$
$$\left. + \frac{(\bar{\imath}_1 e^{j\omega t})^3}{i_0^2\alpha^2}(\Upsilon_3 - e^{-\alpha})^2[5(\Upsilon_3 - e^{-\alpha}) - 3\alpha e^{-\alpha}] \right\} \tag{89}$$

which gives on averaging with respect to x

$$i_e = \bar{\imath}_0 + \bar{\imath}_1 e^{j\omega t}\Upsilon_4 + \left\{ \bar{\imath}_2\Upsilon_4(2\alpha)e^{2j\omega t} + \frac{3\bar{\imath}_1^2}{i_0\alpha}[\Upsilon_3(2\alpha) - \Upsilon_3^2(\alpha)](e^{j\omega t})^2 \right\}$$
$$+ \left\{ \bar{\imath}_3\Upsilon_4(3\alpha)e^{3j\omega t} + \frac{9\bar{\imath}_2\bar{\imath}_1}{2i_0\alpha}[\Upsilon_3(3\alpha) - \Upsilon_3(2\alpha)\Upsilon_3]e^{2j\omega t}\cdot e^{j\omega t} \right.$$
$$\left. + \frac{2\delta i_0\bar{\imath}_1}{3i_0}\left[\Upsilon_1 - \frac{3(\Upsilon_3 - e^{-\alpha})}{\alpha}\right]e^{j\omega t} - \frac{3(\bar{\imath}_1 e^{j\omega t})^3(\Upsilon_3 - e^{\alpha})^3}{i_0^2\alpha^2} \right\} \tag{90}$$

in which all integrations may be confirmed by multiplying (90) through by α^3, differentiating with respect to α, dividing by $3\alpha^2$ and comparing with (89). It should be noted that the multiple-angle vectors $(e^{j\omega t})^2$, $e^{2j\omega t}\cdot e^{j\omega t}$, and $(e^{j\omega t})^3$ are all operated upon by j in both (89) and (90) due to the fact that the transit-angle functions multiplying these vectors reduce to α (times a constant) at low frequencies. We do not, however, need to interpret i_c and i_e in real notation, as it is the potential which we are seeking, and when the potential is obtained it will be found that all functions are in a "nonoperative" condition. As in the case of (76) we obtain $C_0(dv/dt)$ by subtracting the expression for i_e from the total current given by (75a), and integrate with respect to time. We then obtain (compare (78))

$$C_0 v = \frac{\bar{\imath}_0\tau_0}{2} + \bar{\imath}_1\frac{\alpha}{2}\Upsilon_6\frac{e^{j\omega t}}{j\omega} + \left[\bar{\imath}_2\alpha\Upsilon_6(2\alpha)\frac{e^{2j\omega t}}{2j\omega} - \frac{\bar{\imath}_1^2\alpha\Upsilon_{11}}{6i_0}\frac{(e^{j\omega t})^2}{2j\omega}\right]$$
$$+ \bar{\imath}_3\frac{3\alpha\Upsilon_6(3\alpha)}{2}\frac{e^{3j\omega t}}{3j\omega} - \frac{\bar{\imath}_2\bar{\imath}_1\alpha\Upsilon_{15}}{2i_0}\frac{e^{2j\omega t}\cdot e^{j\omega t}}{3j\omega}$$
$$+ \frac{\bar{\imath}_1^3\alpha\Upsilon_{17}}{9i_0^2}\frac{(e^{j\omega t})^3}{3j\omega} - \frac{\bar{\imath}_1\delta i_0\alpha\Upsilon_5}{6}\cdot\frac{e^{j\omega t}}{j\omega}. \tag{91}$$

The expression has been subject to a time integration and now the $j\omega$ cancels with the $j\omega$ in α (leaving τ_0). The functions introduced above are as follows:

$$\Upsilon_{11} = \frac{18}{\alpha^2}\left[\Upsilon_3(2\alpha) - \Upsilon_3{}^2(\alpha)\right] \tag{92}$$

$$\Upsilon_{15} = \frac{9}{\alpha^2}\left[\Upsilon_3(3\alpha) - \Upsilon_3(2\alpha)\Upsilon_3(\alpha)\right] \tag{93}$$

$$\left.\begin{aligned}\Upsilon_{17} &= \Upsilon_7{}^3\\ \Upsilon_5 &= \frac{4}{\alpha}(\Upsilon_4 - \Upsilon_7)\end{aligned}\right\} \Upsilon_7 = \frac{3}{\alpha}(\Upsilon_3 - e^{-\alpha}). \tag{94}$$

Equation (91) enables all components of $\bar{\imath}$ to be calculated if the wave form of v is known, and vice versa. Let us first take $\bar{\imath}$ to be known to be free from second and third harmonics. This is the simplest case. Then, $\delta i_0 = 0$, $\bar{\imath}_0 = i_0$ and (91) becomes

$$\frac{4}{3\tau_0}C_0 v = i_0 + \frac{2}{3}\bar{\imath}_1\Upsilon_6 e^{j\omega t} - \frac{\bar{\imath}_1{}^2\Upsilon_{11}}{9i_0}(e^{j\omega t})^2 - \frac{4\bar{\imath}_1{}^3\Upsilon_{17}}{81 i_0{}^2}(e^{j\omega t})^3. \tag{91a}$$

When the rules of (30) and (35) are applied the coefficients on the right of (91) agree at low frequencies with those calculated from the expansion of $i_0(1+(\bar{\imath}_1/i_0)\cos\omega t)^{2/3}$, though nowhere has a relation between v_0 and i_0 been assumed in the analysis. The solution for the potential in the case studied by Sloane and James[57] cannot, however, as they suggest, be inserted in the 3/2 relation $i = kv^{3/2}$ at the frequency and transit angle in question, and the phase angle between i_c and $\bar{\imath}$ in the case $\xi = \pi/2$ is about 60 degrees instead of only about 30 degrees as appears on their diagram (Fig. 6, I and I_0). Similarly, the relative positions of I_d and I are incorrectly shown. Their expression summarizing the result of graphical computations in the space-charge-limited case is, however, correct within the probable errors of graphical integration; i.e.,

$$V = 2.78 + 0.89\cos(\omega t - 25°47') + 0.03\cos(2\omega t + 60°5')$$

whereas from (91a) may be found

$$v = 2.78_3 + 0.890\cos(\omega t - 26°47') + 0.0343\cos(2\omega t + 71°39')$$
$$- 0.0038\cos(3\omega t + 21°) + 0.0113(\cos\omega t + 67°)$$

in which the last term corrects the argument of the fundamental from $-26°47'$ to $-26°6'$, the modulus remaining at 0.890.

[57] R. W. Sloane and E. G. James, "Transit-time effects in diodes in pictorial form," *Jour. I.E.E.* (London), vol. 79, p. 477; September, (1936).

The next case in order of simplicity is that of no harmonics in the input potential. Equation (91) then gives the various total currents

$$\bar{\imath}_2 e^{2j\omega t} = \frac{\bar{\imath}_1{}^2 \Upsilon_{11} e^{2j\omega t}}{12\dot{\imath}_0 \Upsilon_6(2\alpha)} = \dot{\imath}_0 \frac{3\hat{v}_1{}^2 \Upsilon_{11} e^{2j\omega t}}{16 v_0{}^2 \Upsilon_6(2\alpha) \Upsilon_6{}^2(\alpha)}. \tag{95a}$$

For $\hat{v}_2 = 0$

$$\delta \dot{\imath}_0 = \bar{\imath}_0 - \dot{\imath}_0 = \dot{\imath}_0 \frac{3\hat{v}_1{}^2 \Upsilon_9}{16 v_0 \left| \Upsilon_6{}^2 \right|}. \tag{95b}$$

$$\bar{\imath}_3 e^{3j\omega t} = \frac{2\bar{\imath}_1 \dot{\imath}_0{}^{-2}}{3\Upsilon_6(3\alpha)} \left[\frac{\bar{\imath}_2 \dot{\imath}_0 \Upsilon_{15}}{4} - \frac{\bar{\imath}_1{}^2 \Upsilon_{17}}{36} \right] e^{3j\omega t}$$

$$= -\frac{\dot{\imath}_0 \hat{v}_1{}^3}{64 v_0{}^3} \frac{4\Upsilon_{17}\Upsilon_6(2\alpha) - 3\Upsilon_{11}\Upsilon_{15}}{\Upsilon_6(3\alpha)\Upsilon_6(2\alpha)\Upsilon_6{}^3(\alpha)} e^{3j\omega t}. \tag{95c}$$

In arriving at (95a), (95b), and (95c) it has been sufficiently accurate to take

$$\bar{\imath}_1 e^{j\omega t} = \frac{3\hat{v}_1 e^{j\omega t}}{2 v_0 \Upsilon_6(\alpha)} \dot{\imath}_0 \tag{96}$$

whereas the actual value includes, of course, a third-order contribution

$$\bar{\imath}_1 e^{j\omega t} = \frac{3\hat{v}_1 e^{j\omega t}}{2 v_0 \Upsilon_6(\alpha)} \dot{\imath}_0 - \frac{3\dot{\imath}_0 \hat{v}_1{}^3}{64 v_0{}^3} \left\{ \frac{(4\Upsilon_{17} - 2\Upsilon_9\Upsilon_5)\Upsilon_6(2\alpha) - \Upsilon_{11}\Upsilon_{15}}{\Upsilon_6(2\alpha)\Upsilon_6{}^4(\alpha)} \right\}^* e^{j\omega t} \tag{97}$$

in which the term in Υ_5 comes from the last term of (91).

It will be seen that the impedance functions Υ_6 arrange themselves in the denominator in similar fashion to the impedances $(r_p + z_{n'})$, $n' = 1, 2, 3$, in Section IV, equations (15), which at first sight refers to an entirely different problem. This similarity of analytical form is well brought out by working in closed form; had we contented ourselves with power-series solutions the parallel would, in all probability, have been lost. It is thought that further investigation would enable us to write (15) corrected for transit angle but this will not be attempted here. The equations for comparison with (15) are (96), (95a), and (95c), respectively. Equation (97) compares with the second part of equation (15a). It will be noticed that Υ_9 (see Appendixes II and III) is the appropriate form of Υ_{11} for use with $\delta \dot{\imath}_0$. Similarly, for the third-order contribution to the fundamental, the functions must be arranged in a special way. This is the significance of the asterisk in (97). Some notes on this question will be found in Appendix IV.

The functions Υ_6, Υ_{11}, and Υ_{17} will be found on Fig. 3, which gives, in terms of low-frequency values, the potential harmonics correspond-

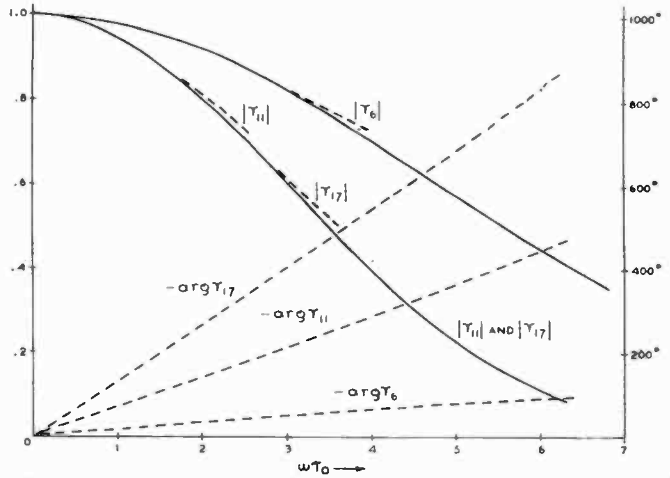


Fig. 3—Modulus (relative to low-frequency value) and argument (right-hand scale) of fundamental (Υ_6), second- (Υ_{11}), and third-harmonic (Υ_{17}) of potential across space-charge-limited plane diode passing a total current $i_0+i_1\cos\omega t$. Broken curves (Υ) show where three terms of series forms (98) give sufficient accuracy.

ing to a total current $i_0+\bar{i}_1\cos\omega t$. The broken lines proceeding from the $|\Upsilon|$ curves indicate the value of ξ for which the series expansions

$$\left.\begin{aligned} |\Upsilon_6| &= \left(1-\frac{13}{300}\xi^2+\frac{11}{12600}\xi^4-\cdots\right)^{1/2}\\ |\Upsilon_{11}| &= \left(1-\frac{11}{100}\xi^2+\frac{9}{1600}\xi^4-\cdots\right)^{1/2}\\ \Upsilon_{17}[=|\Upsilon_7{}^3|] &= \left(1-\frac{3}{80}\xi^2+\frac{1}{1400}\xi^4-\cdots\right)^{3/2}\end{aligned}\right\} \tag{98}$$

no longer give accurate results, though a greater range of ξ may be covered by the series forms if more terms are evaluated. The value of $\arg\Upsilon(\alpha)$ may be obtained, to a greater degree of accuracy, than the series expressions for $|\Upsilon|$, from

$$\left.\begin{aligned} \arg\Upsilon_6 &= -\frac{3}{10}\xi\left(1-\frac{19}{6300}\xi^2+\cdots\right)\\ \arg\Upsilon_{11} &= -\frac{6}{5}\xi\left(1+\frac{17}{10500}\xi^2+\cdots\right)\\ \arg\Upsilon_{17} &= -\frac{9}{4}\xi\left(1+\frac{1}{720}\xi^2+\cdots\right)\end{aligned}\right\}. \tag{99}$$

It may be shown that the above expansions for arg Υ apply over the range $0<\xi<6$ to an accuracy of about 1 per cent as compared with the accurate expressions, obtainable from

$$\left.\begin{aligned}
\Upsilon_6 &= \Upsilon_6'+j\Upsilon_6''\\
&=\frac{12}{\xi^4}\left[2(1-\cos\xi)-\xi\sin\xi\right]+\frac{12j}{\xi^4}\left[2\sin\xi-\xi(1+\cos\xi)-\frac{\xi^3}{6}\right]\\
\Upsilon_{11} &= \Upsilon_{11}'+j\Upsilon_{11}''\\
&=\frac{9}{\xi^6}\left[16(\cos^2\xi-\cos\xi)+16\xi(\sin 2\xi-\sin\xi)\right.\\
&\qquad\left.+\xi^2(1-9\cos 2\xi)-2\xi^3\sin 2\xi\right]\\
&\quad+\frac{9j}{\xi^6}\left[-2\xi^3\cos 2\xi+9\xi^2\sin 2\xi+16\xi(\cos 2\xi-\cos\xi)\right.\\
&\qquad\left.+8(2\sin\xi-\sin 2\xi)\right]\\
\Upsilon_{17} &= \Upsilon_{17}'+j\Upsilon_{17}''\\
&=\frac{27}{\xi^9}\left\{(2\xi\cos\xi-2\sin\xi+\xi^2\sin\xi)\right.\\
&\qquad\left.+j(\xi^2\cos\xi-2\xi\sin\xi+2-2\cos\xi)\right\}^3
\end{aligned}\right\}\tag{100}$$

in which no attempt has been made to rationalize Υ_{17} further by cubing out. Equations (100) will indicate how cumbersome the problem would be if we worked in real notation throughout; equations (65), (92), and (94) look simple in comparison. It should be mentioned that Υ_{11} has been worked out by real methods and agrees with its value in (100). It is perhaps worth writing down the moduli in real notation in the case of Υ_6 and Υ_{17}, as follows:

$$\left.\begin{aligned}
|\Upsilon_6| &= \left[\frac{4}{\xi^8}\left\{\xi^6+12\xi^4(1+\cos\xi)-24\xi^3\sin\xi\right.\right.\\
&\qquad\left.\left.+72\xi^2(1+\cos\xi)-288\xi\sin\xi+288(1-\cos\xi)\right\}\right]^{1/2}\\
|\Upsilon_{11}| &= \left[\gamma_{11}'^2+\gamma_{11}''^2\right]^{1/2}\qquad\text{(see Fig. 4.)}\\
|\Upsilon_{17}| &= \left[\frac{36}{\xi^6}\left\{\frac{\xi^4}{4}+\xi^2\cos\xi-2\xi\sin\xi+2(1-\cos\xi)\right\}\right]^{3/2}
\end{aligned}\right\}.\tag{101}$$

In the case where the input potential is of the form $v_0+v_1\cos\omega t$ the resulting current harmonics will be seen to *increase* with frequency at a rate which increases with the order of the harmonic. For the fundamental, second harmonic, and third harmonic, respectively we require

(97), (95a), and (95c) in which the only functions not yet considered are Υ_5 and Υ_{15}. In series form the second of equations (94) and (93) become

$$\Upsilon_5 = 1 - \frac{3\alpha}{5} + \frac{\alpha^2}{5} - \frac{\alpha^3}{21} + \frac{\alpha^4}{112} - \frac{\alpha^5}{720} + \cdots$$

$$= \sum_{m=0}^{\infty} \frac{12(m+1)(m+2)(-\alpha)^m}{(m+4)!} \tag{102}$$

$$\Upsilon_{15} = 1 - \frac{9\alpha}{5} + \frac{7\alpha^2}{4} - \cdots \tag{103}$$

from which it will be noticed that Υ_{15} is very similar to Υ_{17}, the argument of which is $-(9\xi/4)$. It is, therefore, worth writing down Υ_{17}, and also Υ_{11}, in series form

$$\Upsilon_{17} = 1 - \frac{9\alpha}{4} + \frac{207\alpha^2}{80} - \cdots \tag{104}$$

$$\Upsilon_{11} = 1 - \frac{6\alpha}{5} + \frac{93\alpha^2}{120} - \frac{37\alpha^3}{105} + \frac{351\alpha^4}{2800} - \cdots. \tag{105}$$

Equation (104) indicates that the phase angle of the potential harmonics for harmonicless $\bar{\imath}$ increases at a rate which is rather more than directly proportional to the order of the harmonic, but, referring to Fig. 3 it is seen that the phase difference between the second and third harmonic is almost identical with the phase difference between the fundamental and second harmonic at all transit angles. This is because arg Υ_6 is only $-(3/10)\xi$ (see equations (99)). The complete series expansion for the impedance function is

$$\Upsilon_6 = 1 - \frac{3\alpha}{10} + \frac{\alpha^2}{15} - \frac{\alpha^3}{84} + \frac{\alpha^4}{560} - \frac{\alpha^5}{4320} + \cdots$$

$$= \sum_{m=0}^{\infty} \frac{12(m+2)(-\alpha)^m}{(m+4)!} \tag{106}$$

from which $\Upsilon_6(n\alpha)$ may be obtained for use in (97), (94), and (95). North[46] has suggested that formulas phased against the potential are more useful than those phased against the total current, but his equations ((11), p. 131) are neither one thing nor the other. True, they may bear the correct phase relationship to the potential but the relationship of the modulus is not corrected until the expressions are multiplied by North's F. In view of remarks on p. 132, however, Fig. 3, p. 133, is perfectly in order and conveys a considerable amount of information per square centimeter of graph paper. In my opinion, it is less important to work out Υ_3/Υ_6, $\Upsilon_4/\Upsilon_6 \cdots$ as standard formulas when it is considered that the appearance of Υ_6 in the denominator will be important in connection with tubes working into a load, when these ratios may

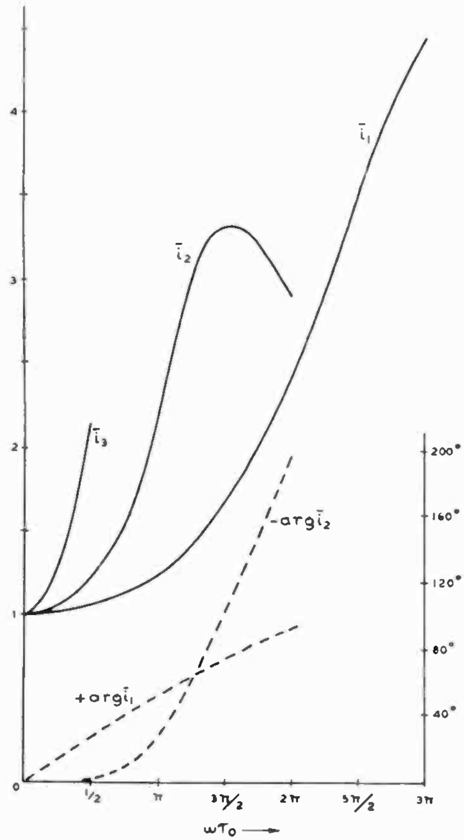


Fig. 4—Modulus and argument of total currents as functions of transit angle for input potential $v_0+v_1 \cos \omega t$. $\bar{i}_1$, $\bar{i}_2$, and $\bar{i}_3$ are referred to their low-frequency values as unity. Arg i_3 (not marked) is zero for $0<\alpha<\pi/4$.

well take the form $r_p\Upsilon_3/(r_p\Upsilon_6+z_1)$, $r_p\Upsilon_4/(r_p\Upsilon_6+z_1)$. These possible forms are quoted by way of stressing the fact that unless we preserve the identity of Υ_6 the physics becomes obscured, but it is not suggested that the matter would necessarily be quite so simple.

In Fig. 4 the various harmonic total currents are shown in terms of their low-frequency values as unity. The increase with frequency at constant $\hat{v}_1$ is, perhaps, more pronounced than would be anticipated from Fig. 3, and is accounted for partly by the existence of multiple-angle impedance factors in the denominators of $\bar{\imath}_2$, $\bar{\imath}_3$. These factors bring it about that the third harmonic of total current increases more rapidly than the second, despite the fact that in the case where the total current is constant and free from harmonics the difference between the frequency characteristic of the modulus of the second and of the third *potential* harmonic is negligible (the points lying for all practical purposes on a common curve though entirely different expressions apply).

It will be noticed that the phase angle of harmonic total currents, referred to input potential, is zero for small transit angles.

The other factor which accounts for the rapid use of the third-harmonic total current is the rapid variation of the quantity Υ_{18}, given by

$$\Upsilon_{18} = 4\Upsilon_{17} - \frac{3\Upsilon_{11}\Upsilon_{15}}{\Upsilon_6(2\alpha)} \doteqdot 1 - \frac{9\alpha}{5} + \frac{283}{200}\alpha^2 \doteqdot \left(1 + \frac{41}{200}\xi^2\right)e^{-j9/5\xi}. \quad (107)$$

Moreover,

$$\Upsilon_6(3\alpha)\Upsilon_6{}^3(\alpha) \doteqdot 1 - \frac{9\alpha}{5} + \cdots \doteqdot \left(1 - \frac{52}{200}\xi^2\right)e^{-j9/5\xi} \quad (108)$$

so that the third harmonic is at low frequencies proportional to the factor $(1+(93\xi^2/200))e^0$, as compared with $(1+(15/200)\xi^2)e^0$ for the second harmonic and $(1+(13/300)\xi^2)e^{+j3/10\xi}$ for the fundamental.

XI. Electron Damping by Harmonics in Receiving Tubes

If the input signal is free from harmonics apart from those produced by the tube, there will be harmonics in the total current between the cathode and the first grid; these harmonics will be given roughly by the above expressions, for indirectly heated tubes approximating to a plane geometry. Now these current harmonics divide, the division between plate current and grid current determining the harmonic currents flowing in the external circuit, or grid input circuit. In order to determine the harmonics in the grid current it is necessary first to obtain an expression for the electron convection current, as the whole of this current passes through the grid, assuming adequate grid bias. When the electron convection current is subtracted from the total current, the difference gives the "hot" displacement current and this current will go mainly to the grid. The "cold" displacement current now contains some harmonics, since the input potential will be altered in wave form owing to harmonics (initially only due to electron displacement current, i.e., the difference between hot and cold displacement currents flowing back through the input circuit. Once the input potential wave form is distorted we may say that both components of the hot displacement current contain harmonics. Before considering the consequence of these remarks, attention is called to the following expression[58] giving the electron convection current at the plane of the first grid of a receiving tube assuming the resultant input potential (after allowing for harmonics) is still free from harmonics; (i.e., the original signal was distorted by just the amount necessary to compensate for harmonics in the grid current of the tube).

$$(i_c)_{g_1}=\frac{2.34\times 10^{-6}}{d'^2}\left(E_{g_1}+\frac{E_p}{\mu}\right)^{3/2}$$

$$\left[\begin{aligned} & 1+\frac{3\hat{v}_1^{\,2}}{16v_0^{\,2}}\left|\Upsilon_{\delta i}\right| \\ & +\left\{\frac{3\hat{v}_1}{2v_0\Upsilon_6}-\frac{3\hat{v}_1^{\,3}}{64v_0^{\,3}}\,\frac{(4\Upsilon_{17}-2\left|\Upsilon_{21}\right|\Upsilon_5)\Upsilon_6(2\alpha)-\Upsilon_{11}\Upsilon_{15}}{\Upsilon_6(2\alpha)\Upsilon_6^{\,4}}\right\}e^{j\omega t}\Upsilon_3 \\ & +\frac{3\hat{v}_1^{\,2}}{16v_0^{\,2}}\left[\frac{\Upsilon_3(2\alpha)\Upsilon_{11}+\frac{4}{3}\alpha\Upsilon_7^{\,2}\Upsilon_6(2\alpha)}{\Upsilon_6(2\alpha)\Upsilon_6^{\,2}}\right]e^{2j\omega t} \\ & -\frac{\hat{v}_1^{\,3}}{64v_0^{\,3}}\left[\frac{4\Upsilon_{17}\Upsilon_6(2\alpha)-3\Upsilon_{11}\Upsilon_{15}}{\Upsilon_6(3\alpha)\Upsilon_6(2\alpha)\Upsilon_6^{\,3}}\right]e^{3j\omega t}\Upsilon_3(3\alpha) \\ & +\frac{\hat{v}_1^{\,3}}{16v_0^{\,3}}\left[\frac{3\Upsilon_{11}\Upsilon_7(2\alpha)e^{2j\omega t}+\left|\Upsilon_9\right|\Upsilon_6(2\alpha)}{\Upsilon_6(2\alpha)\Upsilon_6^{\,3}(\alpha)}\right]\alpha\Upsilon_7e^{j\omega t} \\ & +\frac{\hat{v}_1^{\,3}\alpha\Upsilon_7^{\,2}(5\Upsilon_7-9e^{-\alpha})}{8v_0^{\,3}\Upsilon_6^{\,3}}(e^{j\omega t})^3 \end{aligned}\right] \tag{109}$$

where d' has the significance explained in Section I. In considering the value of Υ_{d_i} it is to be noted that in (89) $\bar{\imath}_0$ includes δi_0. There is definitely no contribution to δi_0 from the bracketed second-order terms of (89). We should thus expect $\Upsilon_{d_{i0}}$ to be as given in (95b) and in Appendix II. The above relation may be obtained by substituting for $\bar{\imath}_1$, $\bar{\imath}_2$, and $\bar{\imath}_3$ from (97), (95a), and (95c), using (96) for $\bar{\imath}_1$ in second- and third-order terms, in (89) for i_c. Equation (109) may be regarded as expressing the condition that a large input signal (not swinging beyond cutoff, however) shall be a pure sine wave at the input to the tube, and the output current would then contain harmonics which, for small values of the ratio $\tau_{g_1g_2}/\tau_{kg_1}$, would be obtained by writing t in (109) as $(t-(\alpha_{g_1g_2}/j\omega))$,

[58] Appendix II deals with the interpretation of such terms of (109) as are likely to cause difficulty.

and, of course allowing for any electrons captured by g_2. At low frequencies the condition reduces simply to

$$(i_c)_{g_1} = \frac{2.34 \times 10^{-6}}{d'^2}\left(E_{g_1} + \frac{E_p}{\mu}\right)^{3/2}\left[1 + \frac{3\hat{v}_1^2}{16v_0^2}(1 + \cos 2\omega t) + \frac{3\hat{v}_1}{2v_0}\cos \omega t - \frac{\hat{v}_1^3}{64v_0^3}(\cos 3\omega t + 3\cos \omega t)\right]. \quad (109a)$$

Also of importance in practice is the condition $(i_c)g_1$ free from harmonics. This, again, may readily be obtained. Referring to (89), the condition is seen to be met if the expressions in the braces { } each vanish, giving the two conditions

$$\left.\begin{aligned} \bar{\imath}_2 e^{2j\omega t} &= \frac{\bar{\imath}_1^2}{i_0\alpha}\,\frac{(\Upsilon_3 - e^{-\alpha})^2}{\Upsilon_3(2\alpha)}\,e^{2j\omega t} = \frac{\alpha\Upsilon_7^2\bar{\imath}_1^2 e^{2j\omega t}}{9\Upsilon_3(2\alpha)i_0} \\ \bar{\imath}_3 e^{3j\omega t} &= \cdots \end{aligned}\right\}. \quad (110)$$

From (91) we have for the input potential, to the second order, subject to (110)

$$v = \frac{\tau_0}{2C_0}\left[i_0 + \bar{\imath}_1\Upsilon_6 e^{j\omega t} + \frac{\bar{\imath}_1^2 e^{2j\omega t}(4\alpha\Upsilon_7^2\Upsilon_6(2\alpha) - 3\Upsilon_{11}\Upsilon_3(2\alpha))}{36 i_0\Upsilon_3(2\alpha)}\right] \quad (111)$$

in which $(e^{j\omega t})^2$ has been split into $(1/2)e^{2j\omega t}$ and 1/2, the direct-current contribution canceling with δi_0, leaving i_0 instead of $\bar{\imath}_0$. Equation (111) gives the input potential required to give no second harmonic in the electron convection current. In case the third harmonic of i_c is also zero there is of course a further set of third-order terms in the expression for v. It must be emphasized that it is not the harmonics *per se* that are of importance so much as the criterion they give as to whether we must expect intermodulation products, including the important case of cross modulation. We can tell from the magnitude of the third harmonic whether cross modulation is important.

Coming now to the question of electron damping, the analysis of D. O. North[46] may be regarded as giving a very close approximation to the first-order grid-to-ground admittance. The third-order contribution to the fundamental would in practical cases be far too small to affect the accuracy of North's expressions. The various assumptions[44] set out by Ferris still apply to the following analysis with the exception of (6); we are now going to allow the input signal to be sufficiently large to give rise to appreciable second harmonic. Third harmonic will not be considered.

Let us take the hot displacement current to create a second-order component of input potential, given by v_2 where

$$C_0 \frac{dv_2}{dt} = \bar{\imath}_2 - \frac{1}{d}\int_0^d (i_c)_2 dx. \tag{112}$$

Existing at the input terminals of the tube there is this potential difference, produced by the flow of hot displacement current in the external circuit. Suppose we knew nothing as to how this second-order current was produced but had means of measuring the phase difference between this current and v_2; and suppose that this information tells us that the current has flown through an impedance containing a resistive component which cannot be attributed to the external circuit. We should then be able to calculate the unknown impedance through which the current passed in the "unknown" part of the circuit. In this way we may *define* the second-order impedance of the cathode-grid path as

$$(z_2)_{kg_1} = \frac{v_2}{\bar{\imath}_2 - (i_c)_2} \tag{113}$$

in which $\bar{\imath} - i_c$ represents the "hot" displacement current. This apparent impedance differs entirely[52] from the input impedance to signals of double frequency, which can of course be obtained from the impedance to fundamental frequency.

We may replace (d/dt) by $2j\omega$ in (112), and then eliminate v_2 by means of (113) giving

$$(z_2)_{kg_1} = \frac{\bar{\imath}_2 - \frac{1}{d}\int_0^d (\bar{\imath}_c)_2 dx}{2j\omega C_0 [\bar{\imath}_2 - (i_c)_2]}. \tag{114}$$

It is preferable to work with the admittance $(y_2)_{kg_1}$. When transit-angle functions are inserted we then obtain, after allowing $j(e^{j\omega t})^2 = (j/2)e^{2j\omega t}$

$$(y_2)_{kg_1} = \frac{2j\omega C_0 \left[1 - \Upsilon_3(2\alpha) + \frac{\bar{\imath}_1^2}{i_0 \bar{\imath}_2} \frac{(\Upsilon_3 - e^{-\alpha})^2}{\alpha}\right]}{\left[1 - \Upsilon_4(2\alpha) + \frac{3\bar{\imath}_1^2}{2 i_0 \bar{\imath}_2}\left\{\frac{\Upsilon_3(2\alpha) - \Upsilon_3}{\alpha}\right\}\right]}. \tag{115}$$

Equation (115) can be dealt with quantitatively if some assumption is made as to the ratio $\bar{\imath}_1/\bar{\imath}_2$. It is convenient to take

$$\frac{\bar{\imath}_2}{\bar{\imath}_1} = \frac{\delta}{\pi} \tag{116}$$

where δ is the decrement of the input tuned circuit, though the approximation involved may be considerable (a) owing to the fact that

$\bar{\imath}_2/\bar{\imath}_1$ are currents in the tube rather than in the tuned circuit and (b) δ includes first-order electron damping, which, though calculable, may affect the relationship (116). In this way preliminary calculation gives that the second-harmonic damping begins to be of importance for input signals greater than about 0.2 times the lumped voltage, but that the second-order capacitative impedance is normal $(c=(4/3)C_0)$ for values of input signal at least up to 0.5 times the lumped voltage. Hence at high frequencies we appear to have a "natural" automatic volume control which will tend to damp down short signals by virtue of second-order electron damping. Ferris states on p. 90 that "the impedance changes by no readily measurable amount with any signal within reason." The present analysis leads one to inquire deeper into this point, but further comment is reserved for the present. Natural automatic volume control is particularly marked in Bakker and de Vries negative anode tetrode,[38] probably, however partly for other reasons.

XII. Effect of Finite Electron-Transit Time on Modulation Products

As pointed out at the beginning of Section IX, it is not necessary to restrict attention to a single input frequency or to simple multiples of a single frequency.

If the circuit arrangements are such that the total current is free from harmonics, the primary equation in the case of three input signals is

$$\ddot{x} = 2.00_2 \times 10^{28}\left[\bar{\imath}_0 + \bar{\imath}_\omega e^{j\omega t} + \bar{\imath}_\Omega e^{j\Omega t} + \bar{\imath}_p e^{jpt}\right]$$

whence without difficulty is obtained the electron convection current

$$i_c = \bar{\imath}_0 + \bar{\imath}_\omega \Upsilon_\omega e^{j\omega t} + \bar{\imath}_\Omega \Upsilon_\Omega e^{j\omega t} + \bar{\imath}_p \Upsilon_p e^{jpt} \tag{117}$$

where,

$$\Upsilon_\Omega \equiv \Upsilon_3(j\Omega\tau).$$

Since τ is the instantaneous transit time, this is allowed for as before, and we obtain, in terms of the direct-current transit time,

$$i_c = \Bigg[\bar{\imath}_0 + \sum \bar{\imath}_\omega \Upsilon_\omega e^{j\omega t} + \sum \frac{2\alpha}{9}\,\frac{\bar{\imath}_\omega{}^2}{\bar{\imath}_0}\,\Upsilon_{14}(e^{j\omega t})^2$$
$$+ \sum \frac{2}{\alpha^2}(\Upsilon_3 - e^{-\alpha})\left\{(\Upsilon_3 - e^{-\alpha})^2 - \alpha e^{-\alpha}(\Upsilon_3 - e^{-\alpha})\right\}\left(\frac{\bar{\imath}_\omega}{\bar{\imath}_0}\right)^3 (e^{j\omega t})^3\Bigg] \tag{118}$$

when $\sum$ indicates that summation is taken over the three frequencies ω, Ω, and p, the two terms not shown being obtained simply by replacing ω by Ω, p respectively (α will change accordingly).

Modulation products of the type $|M_{\omega\Omega p}|\,[\cos(\omega\pm\Omega\pm p)t-\arg M_{\omega\Omega p}]$ will arise by superposition of the various terms, where $M_{\omega\Omega p}$ is a function of all three transit angles; also double frequencies will be involved, but these are usually of less importance. The function $M_{\omega\Omega p}$ could be calculated in any desired case.

If the input potential, rather than the total current, is free from harmonics the problem is exceedingly complex. For we now have to assure for each of the terms in the primary equation an infinite series

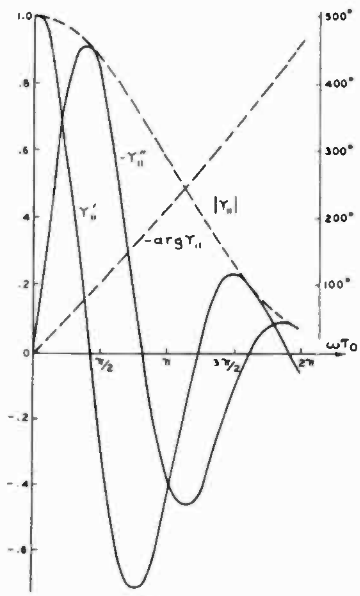


Fig. 5

$$\ddot{x} = 2.00_2 \times 10^{28}\left[\bar{\imath}_0 + \sum_{l=1}^{\infty}\bar{\imath}_l\epsilon^{lj\omega t} + \sum_{m=1}^{\infty}\bar{\imath}_m e^{mj\Omega t} + \sum_{n=1}^{\infty}\bar{\imath}_n e^{njpt}\right]$$

giving

$$i_c = \bar{\imath}_0 + \sum_{l=1}^{\infty}\bar{\imath}_l\Upsilon_{\omega l}e_e^{j\omega t} + \sum_{m=1}^{\infty}\bar{\imath}_m\Upsilon_{\Omega m}e^{mj\Omega t} + \sum_{n=1}^{\infty}\bar{\imath}_n\Upsilon_{pn}e^{njpt} \qquad (119)$$

where,

$$\Upsilon = \Upsilon_3.$$

After expressing i_c in terms of the direct-current transit angle, the procedure is then to obtain the induced current i_e by averaging, and the potential by taking $\bar{\imath}-i_c$ and integrating with respect to t, and then expressing the condition that the potential shall only contain terms in $e^{j\omega t}$, $e^{j\Omega t}$, e^{jpt} and no double- or higher-frequency terms. This stipulation leads to values for the various total currents $\bar{i}_l$, $\bar{i}_m$, $\bar{i}_n$ for all integral values of l, m, and n.

In many cases the frequency characteristic of the various modulation products could be estimated to sufficient accuracy from the harmonic curves of Figs. 3, 4, and 5, making reasonable assumptions to cover the fact that more than one transit angle is involved.

Acknowledgment

It is a pleasure to record my thanks to all workers in the field of high-frequency electronics who have from time to time written to me or sent reprints of their papers. These have been a continual source of encouragement and interest. My grateful acknowledgments are due particularly to those with whom I have had discussions concerning the problems of this paper.

Appendix I. Fundamentals of the Solid-Anode Magnetron

Before coming to consider the solution to the plane magnetron problem, handled by Johannes Müller[59] and myself,[60] it is of interest to note that the cylindrical case has been treated by F. B. Pidduck.[61] As far as I am aware no other papers than these three include the effects of space charge and of magnetic field in a high-frequency treatment including finite electron-transit time, but so many papers have appeared that it is impossible to be quite certain. Thus, many treatments which start by saying "neglecting space-charge forces" are not really neglecting space-charge forces if they afterwards remember to include all effects of electrostatic induction; they are neglecting only the influence of space charge on the electric field, a different matter.

Now Pidduck uses the Eulerian method and after making several transformations obtains a solution for the path under direct-current conditions, and finds that the time taken by electrons to pass from the filament to the turning point is $2.85/107H$ seconds, the greatest radius ρ of sheath which will catch the electrons being $15{,}000\ I_0^{1/2}/H^{3/2}$. The first result corresponds to $\lambda = 17{,}100/H$ on the Okabe basis ($\lambda = 2c\tau_0$), or to $\lambda = 8550/H$ on the theory that oscillations are determined as a result of a single journey. If we let $\omega_0 = eH/m$ the result also corre-

[59] Johannes Müller, "Untersuchungen über Elektronen-strömungen," *Hochfrequenz., und Elektroakustik*, vol. 46, pp. 145–157; November, (1935).
[60] W. E. Benham, "Electronic theory and the magnetron oscillator," *Proc. Phys. Soc.*, vol. 47, p. 1; January, (1935).
[61] F. B. Pidduck, "Space charges in a magnetic field," *Oxford Quarterly Jour. of Math.*, vol. 7, pp. 27, 201–209; September, (1936); "Oscillations with a single-anode magnetron, *ibid.*, pp. 210–213.

sponds to $\omega_0\tau_0=5.038$. On p. 29 of reference 60 I had inferred the result $\omega_0\tau_0=2\pi$ for cylinders on the basis of a table in which five known transit-time ratios were given for planes and two known ratios for cylinders, these being the same as the corresponding ratios for planes; the three blank places were provisionally considered to be similar to those for the plane case. I would, however, have been delighted to withdraw from the tentative position adopted, were it not for the fact that on a second reading of Pidduck's paper I found that I was unable to agree with his final transformation for the path, in which he states that with y'', y', and y all zero, when $x=0$, the equation (primes denote differentiation with respect to x, $=\omega_0\int_a^r dr/\omega_0$)

$$2(\epsilon + y)(y'' + y - x - \alpha) - y'^2 - y^2 = 0 \tag{a}$$

reduces to the equation

$$YY'' + Y^2 = X \tag{b}$$

where $x=2x$ and $y=4Y^2$. The reason for this lies in the fact that the full equation corresponding to (a) is

$$8(Y^3Y'' + Y^4) - 8XY^2 - 2\epsilon X$$
$$+ \{4\epsilon Y^2 - 4\alpha Y^2 + 8\epsilon Y'^2 + 8\epsilon YY'' - \alpha\epsilon\} = 0 \tag{c}$$

in which all terms in $\{\quad\}$ are negligible on the basis that α vanishes for saturated currents and $\epsilon(=m\omega_0^3a^2/4ecI_0)$ vanishes for an infinitely fine filament $(a=0)$. But the term $2\epsilon X$ also vanishes. True, but Pidduck's equation (b) depends on first dividing out by Y^2. Equation (c) then becomes

$$YY'' + Y^2 - \frac{\epsilon X}{4Y^2} = X.$$

Now the last term on the left-hand side is infinite when $Y=0$, unless ϵX is of order Y^2, when the term becomes indeterminate. A path based on (b) can readily be shown to correspond merely to the relation

$$r\ddot{r} + \frac{\omega_0^2r^2}{4} = 3KT \tag{d}$$

which corresponds to taking Langmuir's β equal to unity for all values of r in the relation

$$r = \left(\frac{8e}{3m}i_0\right)^{1/2}\beta\tau_0^{3/2} \tag{e}$$

an approximation also invoked by me on p. 504 of reference 9 in dealing with the case $H=0$. It is, however, doubtful whether one can justify such an approximation when dealing with the path in the magnetron, and the approximation as introduced by me was for the purposes of a first-order solution where its effect might be expected to be of less importance.

Coming now to Pidduck's equation (14) containing the potential V, and the go and return velocities u_1 and u_2, the independent variables being x and t, Pidduck claims that "no approximations have been made so far." Now the approximation already criticized is not, apparently, involved in (14), but it should be mentioned that initial velocities were tacitly neglected at the outset. Pidduck proceeds "If squares and products of $u_1 \cdots$ can be neglected the equations admit an exponential solution in which u_1 can be replaced by $u_1 e^{ivt}$ where u_1 is now a small function of x and the real part is taken in the physical interpretation $\cdots$" from which point a fourth-order equation is involved. On p. 207 Pidduck appears to be in difficulties at the turning point: "Θ and Θ' would have to vanish at each end otherwise u_1 or u_2 would be infinite." It will be noticed that similar infinities would have been arrived at by me,[60] had I neglected initial velocities, in the plane case. Pidduck concludes that "the outer boundary of the cloud does not move," in conflict with my result (Fig. 4 of reference 60), with his previous statement to the effect that squares and products of u_1 cannot be neglected near a turning point, and with our knowledge of conditions at virtual cathodes where large motions can exist in all cases where the alternating electric field is of importance. A further statement "The mass therefore acts like a self-inductance and cannot form part of a self-maintained circuit" is interesting but not, I fear, likely to be the whole story. For Pidduck provides no statement of any kind in regard to the resistive component and one is led to conclude that it is zero. In the sequel, where the plane case is described in its present form, it will be seen that the equivalent resistance in this case may be negative, and acts in series with an inductance. Now, on the basis of the known solutions[9] for planes and cylinders in the case $H=0$, and of the known solution for planes in the case $H\neq 0$ it is possible to make a reasonable guess as to the cylindrical case in the case $H\neq 0$, and from this it does appear that the resistance will not be negative; at the same time, it will not be zero. These last remarks, it must be remembered, apply to the space-charge region and not to the temperature-limited condition. While the above comments are mainly adverse, it will be realized that the paper as a whole is important. The coexistence of electron streams moving in opposite directions introduces difficulties which, when space-charge forces are included render a solution under alternating-current conditions well-nigh impossible. If Pidduck's attempt fails, it is at the same time apparent that no one else has yet succeeded in this problem. From the aesthetic point of view, the paper is unquestionably the most elegant treatment of transit-time problems that has yet been furnished in Eulerian co-ordinates.

We now come to consider the plane magnetron. Müller[59] showed that, if the magnetic field is tilted, so that the angle between magnetic and electric forces differs from $\pi/2$, the impedance of a plane diode can be expressed

$$Z = Z_{\pi/2} \cos^2 \phi + Z_0 \sin^2 \phi$$

where ϕ is the angle between the plates and the lines of magnetic force. In Müller's expression for Z (equations (44) and (45)) the initial velocities do not appear explicitly, but the reason for this will be apparent when we have studied the general expression. Müller's solution likewise depends on all the electrons reaching the plate, a state of affairs which Müller shows can occur for magnetic fields in excess of the critical value in the case $\phi \neq 0$. Now, it is, in practice, very difficult indeed to arrange that $\phi = 0$ exactly, and Müller's treatment brings out the powerful effect of a very small angle of tilt on the ability of the electrons to reach the anode. Such an effect had been overlooked by me in my conclusion that the electrons would turn back at $\theta = 2\pi$ in the case $\phi = 0$. For Müller's treatment now throws doubt on this conclusion; it is necessary to approach the limit very carefully in order to see what happens. On page 43 of reference 60 I had mentioned that I had obtained a solution for the potential, i.e., for the impedance, in the particular case of initial velocities zero, but such was the importance then attached by me to initial velocities that I considered it to be not worth while writing down the solution. Furthermore, no negative resistance was predicted for $\omega_0\tau_0 = 2\pi$; the negative resistance at $(5/2)\pi$ was considered as of academic interest, as this condition meant that electrons were returning to the cathode. Nevertheless, it is no longer merely of academic interest if by any means the electrons do in fact reach the anode at $(\omega\tau_0 = 5\pi/2)$ except insofar as the solid anode magnetron is, of course, of far less immediate commercial interest than the split-anode device.

Müller's paper is in semicomplex notation, i.e., functions of $\omega_0\tau_0$ are in real notation and functions of $\omega\tau_0$ are in complex notation. This procedure shortens the work considerably, though in the Eulerian treatment it would, perhaps, not have been so useful. It is gratifying to report complete agreement between the solution as obtained by Eulerian and Lagrangian treatments, though owing to a choice of boundary conditions which leads to difficulties in the temperature-limited condition, my earlier solution[60] was restricted to the space-charge-limited region. Müller's solution[59] applies everywhere in the temperature-limited region in the case initial velocities are constant (and independent of t) and in the space-charge region *assuming* $\ddot{x}_0=0$. We now come to consider the most general expression possible in case all electrons are collected.

In semicomplex notation (α imaginary, θ real), we have to define the following symbols, many of which have already appeared.

$Z_{\pi/2}$ = complex impedance between electrodes $x=0$ and $x=d$, in ohms, in case $\phi=0$

$j=\sqrt{-1}$

$\omega=2\pi\times$ frequency

C_0 = electrostatic capacitance in farads

Λ = "electronic constant" $=2.00_2\times10^{28}$ erg centimeters coulomb^{-1} gramme^{-1} as in (70)

i_0 = direct current in amperes

τ_0 = direct-current electron-transit time in seconds between the electrodes

a = effective electrode area in square centimeters

$\Upsilon(\alpha,\theta)$ = transit-angle function; $\Upsilon(0,0)=1$

$\alpha=j\omega\tau_0$

$\theta=\omega_0\tau_0$

$$\left[\omega_0=\frac{e}{m}H,\ H=\text{applied magnetic field; electromagnetic units}\right]$$

$$\Upsilon_1(\alpha,\theta)=\frac{6}{\alpha(\alpha^2+\theta^2)}\left[\alpha-2\left(\frac{\alpha^2}{\alpha^2+\theta^2}\right)+\left\{\alpha\left(\frac{\alpha^2-\theta^2}{\alpha^2+\theta^2}\right)\frac{\sin\theta}{\theta}+2\left(\frac{\alpha^2\cos\theta}{\alpha^2+\theta^2}\right)\right\}e^{-\alpha}\right]$$

$\tilde{\imath}_c(0,t)$ = electron convection-current ripple in amperes at plane $x=0$, at instant t

$\tilde{\imath}_1$ = fundamental total current ripple in amperes at instant t

$\ddot{x}_0$ = direct-current electron acceleration in centimeters per second^{-2} at plane $x=0$

$\dot{x}_0$ = direct-current electron velocity in centimeters per second^{-1} at plane $x=0$

d = distance between electrodes in centimeters.

$$\Upsilon_3(\alpha,\theta)=\frac{2}{\alpha^2+\theta^2}\left[1-e^{-\alpha}\cos\theta-\alpha e^{-\alpha}\frac{\sin\theta}{\theta}\right]$$

$$\Upsilon_1(\alpha,\theta)=\frac{1}{\alpha^2+\theta^2}\left[\alpha(1-e^{-\alpha}\cos\theta)+\theta\sin\theta e^{-\alpha}\right]$$

$\dot{x}'_0$ = electron velocity at plane $x=0$, at instant t.

Then the first-order impedance as described above has the components

$$Z_{\pi/2}=\frac{1}{j\omega C_0}\left[1-\frac{\Lambda \bar{i}_0\tau_0^3}{6ad}\Upsilon_4(\alpha,\,\theta)\right.$$
$$\left.-\frac{\bar{i}_c(0,\,t)}{\bar{i}_1 d}\left\{\frac{\ddot{x}_0\tau_0^2}{2}\Upsilon_3(\alpha,\,\theta)+\dot{x}_0\tau_0\Upsilon_1(\alpha,\,\theta)\right\}-\frac{\bar{i}_0\tau_0^2}{2\bar{i}_1 d}\cdot\frac{\partial\dot{x}_0'}{\partial t}\cdot\Upsilon_3(\alpha,\theta)\right] \quad (120)$$

$Z_0=(Z_{\pi/2})_{\theta=0}=$ impedance in the absence of a magnetic field.

The above may be commented upon as follows: First, the impedance is given in practical units and the currents are actual currents in amperes (not currents per unit area). The area a only appears once. Second, Λ is a universal constant ($\Lambda=(4/10)\pi(e/m)c^2$), apart from relativity changes in m, and has a numerical value which happens to be an integer to within 0.1 per cent. Third, the ratios $(i_c(0,\,t))/\bar{i}$, $(1/\bar{i})\,(\partial\dot{x}_0'/\partial t)$ are independent of t, as must of course be the case. Fourth, Z_0 is exactly equivalent to the expression first given by Llewellyn[11] in a somewhat different form. Z_0 is obtained by writing $\theta=0$ in the Υ functions. Fifth, and this is most important, the electrons leaving the plane $x=0$ must reach the plane $x=d$ for the expression to apply. The values of H for which this is possible depend on the angle ϕ, as pointed out by Müller.

While the procedure is not straightforward in the general case, it is still true[62] that (120) may be applied to any electrode pair, and, if $\dot{x}_0$, $\ddot{x}_0$ refer to the initial components of velocity normal to the electrodes, the expression is applicable to the case where the electrons are initially moving at an angle, as would occur, for example, in the grid-anode space of a triode subject to an applied magnetic field. In this case the electrons would be bent away magnetically from normal (or radial) paths in the cathode-grid space and would enter the grid-anode space at an angle, but having normal *components* of velocity $\dot{x}_0$, $\ddot{x}_0$ which could be calculated by applying the direct-current magnetron theory to the cathode-grid space. Some difficulty attaches to the estimation of $i_c(0,\,t)$, but in the triode case this will refer to the grid plane and may be calculated by applying the alternating-current magnetron theory to the (grid-cathode) space, which gives an expression of the type

[62] Caution is necessary, depending on whether the electrons alight on or simply pass through an electrode. In the former case the impedance is of the type $Z=v/(i_e-i_c)$. If admittances of two successive spaces are considered as in parallel, so that

$$Y_1=\left(\frac{i_e-i_c}{v}\right)_1 \qquad -Y_2=\left(\frac{i_e-i_c}{v}\right)_2$$

and if it also happens that on both sides of the electrodes bounding the two spaces i_c is the same (i.e., the electrode does not collect), the resultant admittance is independent of i_c and is given by $Y=Y_1+Y_2=[(i_e)_1-(i_e)_2]/v$ which is of the type $(1/Z)_1+(1/Z_2)$ where Z is given by (120).

$$i_c(0, t) = i_c(-r, t)e^{-\alpha_1} + \frac{\Lambda(\tau_0)_1^2}{2\dot{x}_0}\left\{\bar{\imath}_1\Upsilon_3(\alpha_1, \theta_1) - i_c(-r, t)e^{-\alpha_1}\frac{(1-\cos\theta_1)}{(\frac{1}{2}\theta_1^2)}\right\} \quad (121)$$

where $i_c(-r, t)$ is the electron convection-current ripple at the potential minimum and may usually (space-charge case) be taken as equal to the first-space total current $\bar{\imath}_1$; (τ_0) is the first-space direct-current transit time. In the temperature-limited case, $i_c(-r, t)$ is taken as zero, so that the formula simplifies still further to

$$i_c(0, t) = \frac{\Lambda(\tau_0)_1^2}{2\dot{x}_0}\bar{\imath}_1\Upsilon_3(\alpha_1, \theta_1) \quad (122)$$

in which $\Upsilon_3(\alpha, \theta)$ as listed above applies strictly to a plane geometry only. It should be mentioned that a term similar to the last term of (120) has been omitted from (121), since it can be shown that for a Maxwellian distribution (and neglecting noise fluctuations) the velocity ripple is zero at the potential minimum, as well as at the actual cathode surface. Müller's solution to the magnetron problem corresponds to placing $i_c(0, t) = 0$, $\dot{x}_0' = 0$ in (120), whereas my 1935 method gives $i_c(0, t) = \bar{\imath}$, $\dot{x}_0' = 0$, corresponding to space-charge limitation. Equation (120) comprises all possible degrees of space charge, with reservations as to possible extra terms in τ_m when a potential minimum is present, and may be regarded as a joint effort of myself and Müller.

While Müller has given some curves for the impedance for values of θ in excess of the normal critical value 2π, a great deal of work remains to be done in order to investigate fully the many aspects of the general solution (120). One particular case of considerable practical interest is the case $\alpha \to 0$. While Müller has treated this case as a resistance in series with a condenser, the writer has also calculated the diode as a resistance in parallel with a condenser, which provides information which the more simple "Serieschaltung" fails to give. The calculation was communicated to Müller, who is in agreement with it. It may be mentioned that in general considerable complication arises in converting the "Serieschaltung" to the "Parallelschaltung," so that the Serieschaltung should, in the ordinary course of events, be calculated numerically before converting. The only case for which this is not feasible is the case $\alpha = 0$, which involves the careful taking of limits and gives the following results, for $\dot{x}_0$, $\ddot{x}_0$ zero,

$$\frac{C_{\pi/2}}{C_0} = \frac{(3\sin\theta - 2\theta\cos\theta - \theta - \frac{1}{2}\theta^2\sin\theta)(\theta - \sin\theta)}{\{2(1-\cos\theta) - \theta\sin\theta\}^2} \quad (123)$$

$$\frac{R_{\pi/2}}{R_0} = \frac{4(1-\cos\theta) - 2\theta\sin\theta}{\theta(\theta - \sin\theta)}. \quad (124)$$

Where $R_0 = \tau_0/2C_0$ and is the (zero-frequency) diode impedance for negligible magnetic field ($\theta = 0$). Equation (123) reduces to $C_{\pi/2} = 3/5C_0$ for $\theta \to 0$. Equations (123) and (124) specify the properties of the plane magnetron diode at frequencies for which the electron-transit time is less than or equal to about one tenth of the time period $2\pi/\omega$. The neglect of initial velocities would be serious in the neighborhood of $\theta = 2\pi$, and the negative infinity shown in Fig. 6 is spurious for this reason. At the same time, it is reasonably certain, on both experimental and theoretical grounds, that a negative value of apparent capacitance

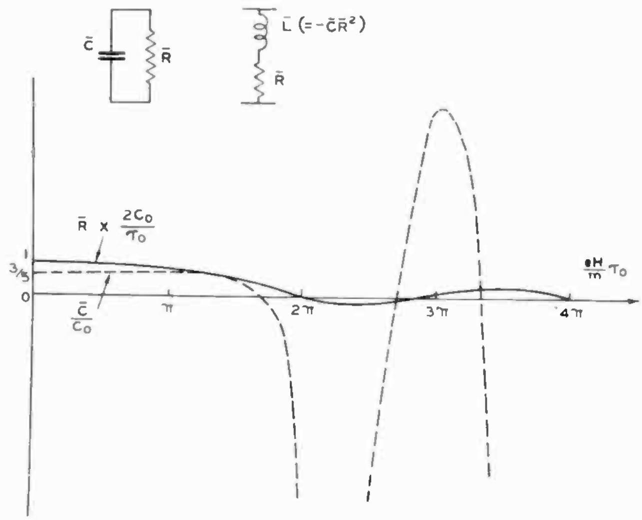


Fig. 6—Plane magnetron as resistance $\overline{R}$ in parallel with capacitance $\overline{C}$ or as resistance $\overline{R}$ in series with inductance $\overline{L}$ is given by $-\overline{C}\overline{R}^2$, for frequencies low compared with $1/\tau_0$.

does arise. It should also be mentioned that since ϕ has been taken as zero (magnetic and electric fields crossed) the critical field is strictly given by $\theta = 2\pi$. For the justification of that part of the diagram to the right of $\theta = 2\pi$ the reader is referred to Müller's paper, which brings out the important fact that even for very small angles of tilt the critical field may be considerably increased.

Fig. 6 predicts maximum negative resistance at $\theta \doteq (5/2)\pi$, a fact which I noticed in 1933, but which was discarded as of academic interest, owing to the opinion that since the critical field occurred for $\theta = 2\pi$, the negative resistance would never be attained. Indeed, at that time I was also somewhat concerned by the fact that the velocity ripple assumed an infinite value when $\theta = 2\pi$. Despite numerous checks on the solution, this physically abhorrent result persisted, and was not cleared up until initial velocities and accelerations were taken into account. In this respect Müller encountered no difficulty as he did not attempt to work with the electron velocity. The existence of an emission velocity can only be taken into account by including such velocity specifically, except in the case of temperature limitation, and in this case the middle term of (120) is zero in any case, since i_c is then zero. The last term of (120) must be taken into account in some cases. While the infinity in Fig. 6 seems to justify the importance attached to the inclusion of initial velocities in the case $\partial v_0/\partial x = 0$ when $x = 0$, it should be emphasized that Müller's treatment including the effect of tilted field provides information in regard to the value of critical field which the inclusion of initial velocities alone does not give. On the other hand, the inclusion of initial velocities assists the direct-current theory of the magnetron hardly at all, but is believed to be all-important for the alternating-current theory in the case where the angle of tilt is in the neighborhood of zero.

Appendix II. The Rectified Current in Diodes at High Frequencies

As indicated elsewhere,[51] the solution for δi_0 may be different from that given by my previous method, also adopted by Tellegen, Jarvis, and Llewellyn. This possibility is of sufficient importance to warrant working out afresh. The "Müller-Llewellyn" method can be adapted to yield a solution for δ_{i0} in a few lines, as indicated below. Now (86b) contains the second-order variation time (previously required only for the third-order solution).

Let us, therefore, start with the equation ($\Lambda = 2.00_2 \times 10^{28}$ dynes coulomb^{-1} gramme^{-1})

$$\ddot{x} = \Lambda(\bar{\imath}_0 + \delta\bar{\imath}_0 + \bar{\imath}_1 e^{j\omega t} + \bar{\imath}_2 e^{2j\omega t})$$

and let us integrate once and pick out the second-order acceleration

$$\dot{x} = \Lambda\left[\delta\bar{\imath}_0\tau_0 + \bar{\imath}_0\tau_2 + \bar{\imath}_1\tau_1 e^{j\omega(t-\tau_0)} + \frac{\bar{\imath}_2 e^{2j\omega t}}{2j\omega}(1 - e^{-2j\omega\tau_0})\right]$$

Writing $\alpha_n = j\omega\tau_n$ as usual, we obtain with the help of (86b)

$$j\omega\dot{x}_2 = \Lambda\left[\frac{2}{3}\alpha_0\delta\bar{\imath}_0 - \bar{\imath}_0\frac{\alpha_1^2}{\alpha_0} + \frac{1}{2}\bar{\imath}_2 e^{2j\omega t}\{1 - \Upsilon_3(2\alpha)\}\right] \qquad (125)$$

in which some very convenient cancellation has taken place. Now at this stage it is possible to see what was done in all previous calculations. As the second harmonic was not under investigation, and as we required mean values in the answer, what we did, in effect, was to write

$$j\omega\bar{x}_2 = \Lambda\left[\frac{2}{3}\alpha_0\delta\bar{\imath}_0 - \bar{\imath}_0\frac{\alpha_1\alpha_1{}^*}{\alpha_0}\right],$$

where the bar denotes "mean value over a cycle." We then obtain on integration with respect to x and expressing the condition that there is no second-order direct-current potential $\bar{v}_2$ (due to the action of the battery in maintaining the direct-current value undisturbed)

$$\delta\bar{\imath}_0 = \frac{\bar{\imath}_1{}^2}{12\bar{\imath}_0}\Upsilon_9 = \frac{\bar{\imath}_1{}^2}{12\bar{\imath}_0}\frac{18}{(-\alpha^2)}(1 - \Upsilon_3\Upsilon_3{}^*). \qquad (126)$$

Completed in proof. Now, if instead of averaging $\alpha_1{}^2$ before integration, we integrate first, with respect to x and take the time average, we obtain

$$\delta\bar{\imath}_0 = \frac{\bar{\imath}_1{}^2}{12\bar{\imath}_0}\frac{18}{(\alpha^2)}\left|\Upsilon_3(2\alpha) - \Upsilon_3{}^2(\alpha).\right. \qquad (126a)$$

Equation (126a) is brought into line with (126) if we change

$\Upsilon_3(2\alpha)$ to $\Upsilon_3(\alpha + (-\alpha)) = 1$ and $\Upsilon_3{}^2(\alpha)$ to $\Upsilon_3(\alpha)\Upsilon_3(-\alpha) = \Upsilon_3\Upsilon_3{}^*$.

Such procedure for δi_0 corresponds, indeed, to normal practice, the modulus of second harmonic being obtained by taking $\sqrt{[\Upsilon_3(2\alpha) - \Upsilon_3{}^2(\alpha)][\Upsilon_3(-2\alpha) - \Upsilon_3{}^2(-\alpha)]}$. As this quantity has already been designated $|\Upsilon_{11}|$ we may use the symbol Υ_{11} equal to the Υ_9 of (126).

In a paper written after this work was submitted, curves and equations based on (Υ_{11}) rather than on Υ_9 were claimed to represent a revised solution for δi_0.[52] The procedure leading to (126) has been checked again in real notation and I now find that it makes no difference whether the time average is effected before or after the x integration. I greatly regret a statement to the contrary, implicating the previous workers mentioned above, in the paper referred to, in which equation in the footnote of p. 473, also equation (1a), p. 475, and curves A' and B' of Fig. 1 are unequivocally withdrawn. The error arose as a result of working in series form in complex notation. This procedure gives the correct result (126) for δi_0 only if the time average is taken before the space integration. This is because, after the space integration a double transit angle occurs in the closed form, all trace of which is lost in the series form, rendering it impossible to replace 2α by $(\alpha+(-\alpha))$ and so obtain Υ_9. Working in real notation gives a correct result in any case, but it should now be clear that complex notation is capable of giving the correct results also. Thus, up to the point where the space integration is taken there are no multiple angles, and the quantity independent of time is correctly given by $\alpha_1\alpha_1{}^*$. In the case of the third-order contribution to the fundamental closed forms are desirable if correct results using complex notation are to be obtained. See Appendix IV. Curves A and B of this diagram are unaffected. A revised set of curves will be published in the *Wireless Engineer*.

Appendix IV. Third-Order Contribution to Fundamental

We have seen in Appendix II that Υ_{11} is interpreted as

$$\frac{18}{\alpha(\pm\alpha)}\left[\Upsilon_3(\alpha\pm\alpha)-\Upsilon_3(\alpha)\Upsilon_3(\pm\alpha)\right]\cdots \tag{127}$$

where the $+$ or $-$ sign is chosen according as whether we are concerned with second harmonic or with the second-order contribution to direct current terms. The interpretation of third-order terms is along similar lines. Thus, consider the term, occurring in the third-order solution,

$$\frac{\Upsilon_{15}\Upsilon_{11}e^{2j\omega t}\cdot e^{j\omega t}}{\Upsilon_6(3\alpha)\Upsilon_6(2\alpha)\Upsilon_6{}^3(\alpha)} = \frac{\frac{18}{2\alpha\cdot\alpha}\left[\Upsilon_3(2\alpha+\alpha)-\Upsilon_3(2\alpha)\Upsilon_3(\alpha)\right]\frac{18}{\alpha\cdot\alpha}\left[\Upsilon_3(\alpha+\alpha)-\Upsilon_3(\alpha)\Upsilon_3(\alpha)\right]e^{2j\omega t}\cdot e^{j\omega t}}{\Upsilon_6(2\alpha+\alpha)\Upsilon_6(2\alpha)\Upsilon_6{}^3(\alpha)}. \tag{128}$$

In (128) Υ_{15} has been expressed in such a way as to bring out the fact that Υ_{15} is closely related to Υ_{11} and is derived therefrom by changing one of the α's to 2α. The identity of 2α and of α is quite clear, so all we have to do in (128) is to replace $e^{j\omega t}$ by $e^{-j\omega t}$; α by $-\alpha$; and the answer is the first-order contribution. We do not need to repeat this process in respect of $e^{2j\omega t}$ and 2α, since (28) of Section VI enables us to dispense with the conjugate function $e^{-j\omega t}$. Now, this replacement of α by $-\alpha$ must only be made where the function of α concerned is intimately associated with $e^{j\omega t}$, which is going to be changed to $e^{-j\omega t}$. Since part of the expression, namely $\left[(\Upsilon_{11}e^{2j\omega t})/\Upsilon_6(2\alpha)\Upsilon_6{}^2(\alpha)\right]$ corresponds definitely to second-harmonic current, δi_0 being already separated out, the only changes we must make are in $\left[(\Upsilon_{15}e^{j\omega t})/\Upsilon_6(2\alpha+\alpha)\Upsilon_6(\alpha)\right]$, the required contribution being, therefore (after dividing by 2),

$$\frac{\bar{\Upsilon}_{15}\Upsilon_{11}e^{2j\omega t}\cdot e^{-j\omega t}}{2\Upsilon_6(2\alpha)\Upsilon_6{}^3(\alpha)\Upsilon_6{}^*(\alpha)} = \frac{\bar{\Upsilon}_{15}\Upsilon_{11}e^{j\omega t}}{2\Upsilon_6(2\alpha)\Upsilon_6{}^2(\alpha)\left|\Upsilon_6{}^2(\alpha)\right|} \tag{129}$$

where,

$$\bar{\Upsilon}_{15} = \frac{9}{-\alpha^2}\left[\Upsilon_3(\alpha)-\Upsilon_3(2\alpha)\Upsilon_3{}^*(\alpha)\right]. \tag{129a}$$

Appendix III—Table of Υ-Functions Applicable to Plane Geometry

The integral functions listed below are compounded of the variable α and of $e^{\pm\alpha}$. The latter being transcendental, the functions are also transcendental. While the radius of convergence is infinity in all cases so far investigated, the large number of terms of the power-series expansions which must be retained for accurate results at large values of $|\alpha|$ render such expansions sometimes less convenient than the closed forms, which are also particularly convenient for $|\alpha| = n(\pi/2)$ where n is any positive integer. In the table, $|\alpha| = \xi = \omega\tau_0$. The conversion of the closed form of Υ to real notation is effected by replacing α by $j\xi$; $e^{\pm\alpha}$ by $\cos\xi \pm j\sin\xi$. Thus Υ_1 becomes $1/\xi\{\sin\xi + j(\cos\xi - 1)\}$. The mode of showing the power series is one which from experience seems to be at once the most convenient and the most useful for purposes of calculation. The closed forms of $|\Upsilon|$ and $\arg\Upsilon$ are clumsy expressions, and it would only be necessary to use them on rare occasions. The power series forms of $|\Upsilon|$ and $\arg\Upsilon$ are therefore given to a degree of accuracy which suffices for most purposes. Contrary to superficial appearance, the series approximations for $\arg\Upsilon$ are more accurate than those for $|\Upsilon|$, although the latter are usually correct to 1 per cent for $0 < \xi < 3$. It is to be noted that $\arg\Upsilon_1 = \arg\Upsilon_4 = -\xi/2$ exactly. It is also of interest that where $|\arg\Upsilon| > \xi/2$ the two terms of the series expansion have the same sign, while for $|\arg\Upsilon| < \xi/2$ the two terms are of opposite sign. The exception occurs in the case of Υ_8 which is in other respects anomalous. Thus in $|\Upsilon_8|$ the signs shown are all opposite to the normal. Blanks have been left in cases of closed forms too complicated for insertion. The first eight Υ functions refer to first-order transit-time solutions; Υ_9 to Υ_{14} inclusive to second-order; Υ_{15} to Υ_{18} inclusive to third-order solutions. It is a matter of interest that $|\Upsilon_3| = |\Upsilon_2|$, which makes $\arg\Upsilon_1 = \frac{1}{2}(\arg\Upsilon_2 + \arg\Upsilon_3)$. In regard to the purely real functions Υ_9, Υ_{10}, and Υ_{13} the series from under $|\Upsilon|$ is shown in round brackets, in order to distinguish at a glance from other cases, which are as a rule shown up as a square-bracketed term raised to the power of 1/2. In the case of Υ_{14} and Υ_{17} the index is 1 and 3/2, respectively, to emphasize the significance in relation to Υ_7. It is convenient and conducive to accuracy to evaluate the square-bracketed terms numerically, afterwards taking the square root, rather than form a binomial expansion, before making numerical calculations. For this reason $|\Upsilon|$ has normally been left in the form $[|\Upsilon^2|]^{1/2}$. Functions like Υ_{12} are often best worked out by evaluating first the constituents. For example, $\arg\Upsilon_{12} = \arg\Upsilon_{11} - 2\arg\Upsilon_6 - \arg\Upsilon_6 2(\alpha)$.

	No. of equation or equations where found in text	Relation to other functions and in particular to Υ_1 (a) (b) (c) (d) (e)	Closed form of Υ in terms of α	Power series of Υ in terms of α	Modulus and (power series forms) $\lvert\Upsilon\rvert$	Argument $\arg\Upsilon$	Physical Significance
Υ_1	53, 54, 55, 85	$\Upsilon_1 = -\frac{1}{2}\int \Upsilon_1 d\alpha = \Upsilon_3 + \frac{\alpha}{6}\Upsilon_4 = 1 - \frac{\alpha}{2}\Upsilon_2 = \frac{1}{2}(\Upsilon_3 + \Upsilon_2)$	$\frac{1}{\alpha}(1 - e^{-\alpha})$	$1 - \frac{\alpha}{2} + \frac{\alpha^2}{6} - \frac{\alpha^3}{24} + \frac{\alpha^4}{120} - \cdots$	$\left[1 - \frac{\xi^2}{12} + \frac{\xi^4}{360} - \frac{\xi^6}{20610} + \cdots\right]^{1/2}$	$-\frac{\xi}{2}$	Appears in complex impedance of plane diode as transit-angle correcting factor to term involving (direct-current) initial velocities. See (85).
Υ_2		$\Upsilon_2 = -\frac{1}{3}\int \Upsilon_2 d\alpha = \frac{2 - \Upsilon_3}{1 + \alpha} = 2\Upsilon_1 - \Upsilon_3 = \Upsilon_3 + \frac{\alpha}{3}\Upsilon_4 = \frac{2}{\alpha}(1 - \Upsilon_1)$	$\frac{2}{\alpha^2}(\alpha - 1 + e^{-\alpha})$	$1 - \frac{\alpha}{3} + \frac{\alpha^2}{12} - \frac{\alpha^3}{60} + \frac{\alpha^4}{360} - \cdots$	$\left[1 - \frac{\xi^2}{18} + \frac{\xi^4}{720} - \frac{\xi^6}{50400} + \cdots\right]^{1/2}$	$-\frac{\xi}{3}\left(1 - \frac{\xi^2}{270} + \cdots\right)$	Transit-angle correcting factor to electron-velocity ripple in extreme temperature-limited condition.
Υ_3	56, 55a ⋯ 64 ⋯ 85	$\Upsilon_3 = \frac{2}{\alpha^2}\int_0^\alpha ze^{-z}dz = e^{-\alpha} + \frac{1}{\alpha^2}\int_0^\alpha z^2e^{-z}dz$	$\frac{2}{\alpha^2}(1 - e^{-\alpha} - \alpha e^{-\alpha})$	$1 - \frac{2\alpha}{3} + \frac{\alpha^2}{4} - \frac{\alpha^3}{15} + \frac{\alpha^4}{72} - \cdots$	$\left[1 - \frac{\xi^2}{18} + \frac{\xi^4}{720} - \frac{\xi^6}{50400} + \cdots\right]^{1/2}$	$-\frac{2\xi}{3}\left(1 + \frac{\xi^2}{540} + \cdots\right)$	There is more than one physical interpretation of this function. In the forms shown, Υ_3 and $\Upsilon_3 - e^{-\alpha}$ may be interpreted as space averages of an impulse starting from $z = 0$ and traveling to $z = \alpha$. Transit-angle correcting factor to electron convection current ripple in space-charge-limited case and appears twice in (85). In temperature-limited case occurs as $((\alpha/2)\Upsilon_3)$.
Υ_4	63, 65 ⋯ 68 ⋯ 85	$\Upsilon_4 = \frac{3}{\alpha^3}\int_0^\alpha z^2\Upsilon_1(z)dz = 1 - \frac{\alpha}{2}\Upsilon_6 = \Upsilon_3 + \frac{\alpha}{6}\Upsilon_5$	$\frac{6}{\alpha^3}[\alpha - 2 + (\alpha + 2)e^{-\alpha}]$	$1 - \frac{\alpha}{2} + \frac{3\alpha^2}{20} - \frac{\alpha^3}{30} + \frac{\alpha^4}{168} - \cdots$	$\left[1 - \frac{\xi^2}{20} + \frac{3\xi^4}{2800} - \frac{\xi^6}{75600} + \cdots\right]^{1/2}$	$-\frac{\xi}{2}$	Transit-angle correcting factor to electron-velocity ripple and induced-current ripple in space-charge-limited diode. In temperature-limited case, it occurs as $((\alpha/3)\Upsilon_4)$.
Υ_5	91, 94, 102	$\Upsilon_5 = \frac{6}{\alpha}(\Upsilon_4 - \Upsilon_3) = \frac{4}{\alpha}(\Upsilon_4 - \Upsilon_7)$	$\frac{36}{\alpha^4}[2\alpha(2e^{-\alpha} - 1) + 2(e^{-\alpha} - 1) + 3\alpha^2e^{-\alpha}]$	$1 - \frac{3\alpha}{5} + \frac{\alpha^2}{5} - \frac{\alpha^3}{21} + \frac{\alpha^4}{112} - \cdots$	$\left[1 - \frac{\xi^2}{25} + \frac{\xi^4}{1400} - \frac{\xi^6}{132,300} + \cdots\right]^{1/2}$	$-\frac{3\xi}{5}\left(1 + \frac{\xi^2}{3150} + \cdots\right)$	Transit-angle correcting factor to electron-displacement current in the space-charge diode, being difference between "hot" and "cold" displacement current.
Υ_6	77 ⋯ 99, 100, 106	$\Upsilon_6 = \frac{2}{\alpha}(1 - \Upsilon_4)$	$\frac{12}{\alpha^4}\left[\frac{\alpha^3}{6} - \alpha + 2 - (\alpha + 2)e^{-\alpha}\right]$	$1 - \frac{3\alpha}{10} + \frac{\alpha^2}{15} - \frac{\alpha^3}{84} + \frac{\alpha^4}{560} - \cdots$	$\left[1 - \frac{13\xi^2}{300} + \frac{11\xi^4}{12600} - \frac{11\xi^6}{1,069,200} + \cdots\right]^{1/2}$	$-\frac{3\xi}{10}\left(1 - \frac{19\xi^2}{6300} + \cdots\right)$	Transit-angle correcting factor to cold displacement current referred to total current. General impedance factor when initial velocities are negligible.
Υ_7	91, 94 ⋯ 110	$\Upsilon_7 = \frac{3}{\alpha}(\Upsilon_3 - e^{-\alpha}) = \frac{(3\Upsilon_3 - \Upsilon_4)}{2}$	$\frac{3}{\alpha^3}\,2 - (\alpha^2 + 2\alpha + 2)e^{-\alpha}]$	$1 - \frac{3\alpha}{4} + \frac{3\alpha^2}{10} - \frac{\alpha^3}{12} + \frac{\alpha^4}{56} - \cdots$	$\left[1 - \frac{3\xi^2}{80} + \frac{\xi^4}{1400} - \cdots\right]^{1/2}$	$-\frac{3\xi}{4}\left(1 + \frac{\xi^2}{720} + \cdots\right)$	Transit-angle correction factor to variation time and to space-charge-density ripple in the space-charge case.
Υ_8		$\Upsilon_8 = \Upsilon_2 - \alpha\Upsilon_3$; $\Upsilon_3 = 2e^{-\alpha} - \Upsilon_8$	$\frac{2}{\alpha^2}[(1 + \alpha + \alpha^2)e^{-\alpha} - 1]$	$1 - \frac{4\alpha}{3} + \frac{3\alpha^2}{4} - \frac{4\alpha^3}{15} + \frac{5\alpha^4}{72} - \cdots$	$\left[1 + \frac{15}{18}\xi^2 - \frac{7\xi^4}{720} + \frac{\xi^6}{5600} - \cdots\right]^{1/2}$	$-\frac{4\xi}{3}\left(1 - \frac{23\xi^2}{405} + \cdots\right)$	Transit-angle correction factor to space-charge density in the extreme temperature-limited condition.
Υ_9	126	$\Upsilon_9 = \frac{18}{(-\alpha^2)}[1 - \Upsilon_2\Upsilon_2^*] = \frac{18}{\xi^2}[1 - \lvert\Upsilon_2^2\rvert] = \bar{\Upsilon}_{11}$	$\frac{72}{\alpha^4}\left[2 - \alpha^2 - \frac{\alpha^4}{4} + (\alpha - 1)e^{\alpha} - (\alpha + 1)e^{-\alpha}\right]$	$1 + \frac{\alpha^2}{40} + \frac{\alpha^4}{2800} + \cdots$	$\left(1 - \frac{\xi^2}{40} + \frac{\xi^4}{2800} - \cdots\right)$		Transit-angle correction factor for rectified current in the space-charge-limited diode, referred to fundamental total current.
Υ_{10}		$\Upsilon_{10} = \frac{\Upsilon_9}{\lvert\Upsilon_6^2\rvert} = \frac{\bar{\Upsilon}_{11}}{\lvert\Upsilon_6^2\rvert}$		$1 - \frac{11\alpha^2}{600} + \frac{39\alpha^4}{140,000} - \cdots$	$\left(1 + \frac{11\xi^2}{600} + \frac{39\xi^4}{140,000} + \cdots\right)$		Transit-angle correcting factor for rectified current in space-charge-limited diode, referred to fundamental input potential. Potential wave form pure with diode cold.
Υ_{11}	91, 92 ⋯ 99, 100 ⋯ 127	$\Upsilon_{11} = \frac{18}{\alpha^2}[\Upsilon_2(2\alpha) - \Upsilon_2^2(\alpha)]$; $\bar{\Upsilon}_{11} = \Upsilon_9$		$1 - \frac{6\alpha}{5} + \frac{93\alpha^2}{120} - \frac{37\alpha^3}{105} + \frac{351\alpha^4}{2800}$	$\left[1 - \frac{11}{100}\xi^2 + \frac{9}{1600}\xi^4 - \cdots\right]^{1/2}$	$-\frac{6}{5}\xi\left(1 + \frac{17\xi^2}{10500} + \cdots\right)$	See Fig. 4. Transit-angle correcting factor of second-harmonic induced current and referred to fundamental total current.
Υ_{12}		$\Upsilon_{12} = \Upsilon_{11} \div [\Upsilon_6(2\alpha)\Upsilon_6^2(\alpha)]$			$\left[1 + \frac{3}{20}\xi^2 + \cdots\right]^{1/2}$	$\frac{577\xi^3}{52500} + \cdots$	See Fig. 5. Transit-angle correcting factor of total current second-harmonic referred to harmonic-free input potential.
Υ_{13}	128, 129	$\Upsilon_{13} = \lvert\Upsilon_{11} \div \Upsilon_6^2(\alpha)\rvert = \frac{\sqrt{\{\Upsilon_2(2\alpha) - \Upsilon_2^2(\alpha)\}\{\Upsilon_2^*(2\alpha) - \Upsilon_2^{*2}(\alpha)\}}}{\Upsilon_6\Upsilon_6^*}$		$1 + \frac{7\alpha^2}{600} + \frac{171\alpha^4}{140,000} + \cdots$	$\left(1 - \frac{7\xi^2}{600} + \frac{171\xi^4}{140,000} - \cdots\right)$		Erroneous transit-angle correcting factor for rectified current in space-charge-limited diode, referred to fundamental input potential (potential wave form pure with diode cold).
Υ_{14}	118	$\Upsilon_{14} = \Upsilon_7^2 = \Upsilon_{17}^{2/3}$		$1 - \frac{3\alpha}{2} + \frac{93\alpha^2}{80} - \frac{37\alpha^3}{60} + \frac{351\alpha^4}{1400}$	$\left(1 - \frac{3}{80}\xi^2 + \frac{1}{1400}\xi^4 - \cdots\right)^1$	$-\frac{3\xi}{2}\left(1 + \frac{\xi^2}{720} + \cdots\right)$	Square of variation-time correcting factor Υ_7.
Υ_{15}	91, 93, 103	$\Upsilon_{15} = \frac{9}{\alpha^2}[\Upsilon_2(3\alpha) - \Upsilon_2(2\alpha)\Upsilon_2(\alpha)]$. For $\bar{\Upsilon}_{15}$ see App. IV		$1 - \frac{9\alpha}{5} + \frac{7\alpha^2}{4} - \cdots$	$\left[1 - \frac{13}{50}\xi^2 + \cdots\right]^{1/2}$	$-\frac{9}{5}\xi(1 + \cdots)$	Υ_{15} and Υ_{16} are the constituents of third-harmonic total current phased against fundamental input potential.
Υ_{16}	108	$\Upsilon_{16} = \Upsilon_6(3\alpha)\Upsilon_6^2(\alpha)$		$1 - \frac{9\alpha}{5} + \frac{47\alpha^2}{25} - \cdots$	$\left[1 - \frac{13}{25}\xi^2 + \cdots\right]^{1/2}$	$-\frac{9}{5}\xi(1 + \cdots)$	
Υ_{17}	91, 94 ⋯ 99, 100	$\Upsilon_{17} = \Upsilon_7^3 = \Upsilon_{14}^{3/2}$. For $\bar{\Upsilon}_{17}$ see App. IV		$1 - \frac{9\alpha}{4} + \frac{207\alpha^2}{80} - \cdots$	$\left[1 - \frac{3}{80}\xi^2 + \frac{1}{1400}\xi^4 - \cdots\right]^{3/2}$	$-\frac{9}{4}\xi\left(1 + \frac{\xi^2}{720} + \cdots\right)$	Cube of variation-time correcting factor Υ_7. Transit-angle correcting factor of third-harmonic induced current and potential referred to fundamental total current.
Υ_{18}	107	$\Upsilon_{18} = 4\Upsilon_{17} - \frac{3\Upsilon_{11}\Upsilon_{15}}{\Upsilon_6(2\alpha)}$		$1 - \frac{9\alpha}{5} + \frac{283\alpha^2}{200} + \cdots$	$\left[1 + \frac{41}{100}\xi^2 + \cdots\right]^{1/2}$	$-\frac{9}{5}\xi(1 + \cdots)$	See Υ_{15}.

Although, as stated above, we do not need the conjugate expression, it is to be noted that in order to obtain it we need to change 2α to -2α (and 2ω to -2ω) in the expression $[(\Upsilon_{15}e^{2j\omega t}\cdot e^{j\omega t})/\Upsilon_6(2\alpha+\alpha)\Upsilon_6]$, while the signs of all α's in $[(\Upsilon_{11}e^{2j\omega t})/\Upsilon_6(2\alpha)\Upsilon_6^2(\alpha)]$ must now be changed. We then obtain as the conjugate to (129)

$$\frac{\overline{\Upsilon}_{15}{}^*\Upsilon_{11}{}^*e^{-j\omega t}}{2\Upsilon_6{}^*(2\alpha)\Upsilon_6{}^{*2}\,|\,\Upsilon_6{}^2\,|} \tag{129*}$$

where $\Upsilon_{11}{}^*=18[\Upsilon_3{}^{*2}-\Upsilon_3{}^*(2\alpha)]/\alpha^2$; $\Upsilon_{15}{}^*=9[\Upsilon_3(\alpha)\Upsilon_3{}^*(2\alpha)-\Upsilon_3{}^*]/\alpha^2$. It is to be particularly noted that the treatment of Υ_{11} is different from that of Υ_{15}, in that the latter contains both (2α) and (α), and consequently changes in both cases, while in the present case Υ_{11} does not contain α at all, only $(2\alpha)/2$ or (2α). The behavior of Υ_{15} in the above example is precisely analogous to that of Υ_{11} in the second-order solutions. Actually this does not mean that $\overline{\Upsilon}_{11}$, obtained by using the minus signs in (127), does not occur in third-order solutions; it does, but not in the term considered above. Indeed, the importance of including a contribution from δi_0 has already been stressed in the text.

We are further interested in the term

$$\frac{\Upsilon_{17}(e^{j\omega t})^3}{\Upsilon_6(3\alpha)\Upsilon_6{}^3(\alpha)}=\frac{\Upsilon_7{}^3(e^{j\omega t})^3}{\Upsilon_6(\alpha+\alpha+\alpha)\Upsilon_6{}^3(\alpha)}\,. \tag{130}$$

There are three ways of changing α to $-\alpha$ (ω to $-\omega$) and only one way of leaving the expression with signs unchanged. Our third-order contribution to the fundamental is thus

$$\frac{3\Upsilon_7{}^*\Upsilon_7{}^2(e^{-j\omega t})(e^{j\omega t})^2}{4\Upsilon_6(\alpha)\Upsilon_6{}^2(\alpha)\Upsilon_6{}^*(\alpha)}=\frac{3\,|\,\Upsilon_{14}\,|\,\Upsilon_7e^{j\omega t}}{4\Upsilon_6{}^2\,|\,\Upsilon_6{}^2\,|}\,. \tag{131}$$

We check the above by taking the conjugate. For this we need to change two α's out of three, at one time,

$$\frac{3\Upsilon_7{}^{*2}\Upsilon_7(e^{-j\omega t})^2e^{j\omega t}}{4\Upsilon_6{}^*(\alpha)\Upsilon_6{}^{*2}(\alpha)\Upsilon_6(\alpha)}=\frac{3\,|\,\Upsilon_{14}\,|\,\Upsilon_7{}^*e^{-j\omega t}}{4\Upsilon_6{}^{*2}\,|\,\Upsilon_6{}^2\,|}\,. \tag{131*}$$

The total third-order contribution to the fundamental, as given in (97), must now be written

$$-\frac{3i_0v_1{}^3}{64v_0{}^3}\left\{\frac{(4\,|\,\Upsilon_{14}\,|\,\Upsilon_7-2\Upsilon_9\Upsilon_5)\Upsilon_6(2\alpha)-\overline{\Upsilon}_{15}\Upsilon_{11}}{\Upsilon_6(2\alpha)\Upsilon_6{}^2\,|\,\Upsilon_6{}^2\,|}\right\}. \tag{132}$$

The bracketed terms of (109) must also be replaced by those of (132).

We have in (109) to consider further third-order contributions to the fundamental. In the last line but one we have the quantity

$$\frac{\Upsilon_{11}\Upsilon_7(2\alpha)e^{2j\omega t}\cdot\alpha\Upsilon_7 e^{j\omega t}}{\Upsilon_6(2\alpha)\Upsilon_6{}^3(\alpha)}. \tag{133}$$

For the same reasons as given in connection with Υ_{15}, when we change α to $-\alpha$, $\Upsilon_{11}/\Upsilon_6(2\alpha)\Upsilon_6{}^2$ remains unchanged. The required third-order contribution in this case is only one third of the third harmonic. With (45a),[36] as guides in case of α small we find

$$\frac{+\ \Upsilon_{11}\Upsilon_7(2\alpha)\alpha\Upsilon_7{}^*e^{j\omega t}}{6\Upsilon_6(2\alpha)\Upsilon_6(\alpha)\mid\Upsilon_6{}^2\mid}. \tag{134}$$

The only remaining term for consideration, see last term of (109), is

$$\frac{\alpha[\Upsilon_7{}^2(5\Upsilon_7 - 9e^{-\alpha})]}{\Upsilon_6{}^3(\alpha)}(e^{j\omega t})^3. \tag{135}$$

We note that this term is, like (133), in an operative condition with (35a) as guide, the third-order contribution to the fundamental is found to be

$$\frac{\Upsilon_7{}^2[5\Upsilon_7{}^* - 9e^{+\alpha}] + 2\mid\Upsilon_{14}\mid(5\Upsilon_7 - 9e^{-\alpha})}{6\Upsilon_6\mid\Upsilon_6{}^2\mid}(+\ \alpha e^{j\omega t}). \tag{136}$$

With the help of Appendix III and of the expansions

$$\bar{\Upsilon}_{15} = 1 - \frac{3}{5}\alpha + \frac{\alpha^2}{4} - \cdots \tag{137}$$

$$\bar{\Upsilon}_{17} = 1 - \frac{3}{4}\alpha + \frac{27\alpha^2}{80} - \cdots \tag{138}$$

the contribution in the brackets of (132) evaluates to $(1+\xi^2/8)e^0$, for small values of ξ, from which it is seen that (after allowing for the minus sign in (132)) this contribution to the fundamental leads the first-order contribution in phase by $(\pi - 3j\xi/10)$ radians.

Thus for $0<\xi<\pi/4$ the first-order term alone has a phase angle $(3j\xi/10)$, which corresponds to capacitance current (diode capacitance $3C_0/5$). The fact that for second- and third-order contributions the phase angle is (for $0<\xi<\pi/4$) unaffected by electron inertia corresponds to the condition $v_2 = v_3 = 0$ holding, not only at the electrodes, but at intermediate points.

Proceedings of the Institute of Radio Engineers
Volume 26, Number 9 *September, 1938*

CHARACTERISTICS OF THE IONOSPHERE AT WASHINGTON, D. C., JULY, 1938*

By

T. R. Gilliland, S. S. Kirby, and N. Smith

(National Bureau of Standards, Washington, D. C.)

DATA on the critical frequencies and virtual heights of the ionosphere layers for July, 1938, are given in Fig. 1. Fig. 2 gives the maximum frequencies which could be used for radio sky-wave communication by way of the normal E, F, F_1, and F_2 layers. The

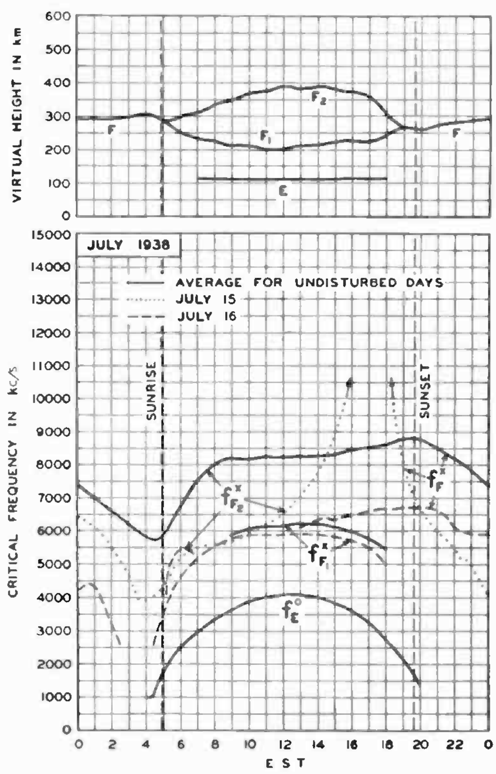


Fig. 1—Virtual heights and critical frequencies of the ionosphere layers, July, 1938. The solid-line graphs represent the average for undisturbed days. The dotted and dashed graphs represent values for the ionosphere-storm days indicated.

* Decimal classification: R113.61. Original manuscript received by the Institute, August 10, 1938. This is one of a series of reports on the characteristics of the ionosphere at Washington, D. C. For earlier publications see Proc. I.R.E.,

flatness of the graphs indicates very little change in the maximum usable frequencies from day to night. The critical frequencies and the maximum usable frequencies were approximately the same as for June.

Because of reflections from clouds of sporadic E layer during frequent irregular periods, the maximum usable frequencies often greatly exceeded the regular, dependable values shown in the graphs. Because of the erratic occurrence of these transmissions they cannot be included in the graphs.

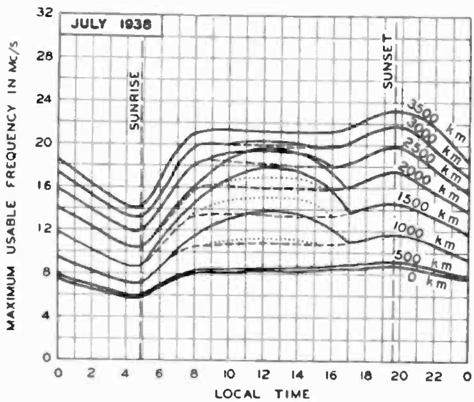


Fig. 2—Maximum usable frequencies for radio sky-wave transmission, July, 1938. Solid-line graphs represent average values for undisturbed days, for dependable transmission by the regular ionosphere layers. The values shown were considerably exceeded during frequent irregular periods by reflections from clouds of sporadic E layer. For distances of 1000, 1500, and 2000 kilometers, the dotted and dashed portions of the graphs represent maximum usable frequencies for F_1- and F_2-layer transmission, respectively, when these were less than those determined by the E layer. For distances of 2500 and 3000 kilometers the dashed line represents maximum usable frequencies for F_2-layer transmission when these were less than those determined by the F_1 layer.

July was marked by considerable ionosphere irregularities, both ionosphere storms and radio fade-outs. A large proportion of the fade-outs occurred within approximately the week preceding the most severe ionosphere storms. This indicates that the two effects are produced by different agencies associated with the same solar source.

The ionosphere storms which occurred during July are listed in Table I.

---

vol. 25, pp. 823–840; July, (1937), and a series of monthly reports beginning in Proc. I.R.E., vol. 25, pp. 1174–1191; September, (1937). Publication approved by the Director of the National Bureau of Standards of the U. S. Department of Commerce.

TABLE I

IONOSPHERE STORMS (APPROXIMATELY IN ORDER OF SEVERITY)

Date and hour, E.S.T.	h_F before sunrise, km	Minimum $f_F{}^x$ during day (before sunrise), kc	Noon $f_{F_2}{}^x$ kc	Magnetic character[1]		Ionosphere character[2]
				00–12 G.M.T.	12–24 G.M.T.	
July 15 after 1200	—	—	6500	1.1	1.6	2
July 16	482	<2500	<5900	1.1	1.1	2
July 17 until 1700	350	4200	6200	0.4	0.1	1
July 30	398	4000	<6000	1.6	0.9	1½
July 31 until 0500	314	4400	—	0.2	0.0	½
July 13 after 1800	—	—	—	0.2	0.9	½
July 14	310	4700	6300	0.6	0.6	1
July 15 until 1200	338	3900	6500	1.1	1.6	1
July 1 after 0400	298	4800	6400	0.7	0.6	1
July 2	310	5300	6100	0.6	0.1	1
July 10	no vertical incidence data			0.9	0.9	—
July 11 until 1400	326	4100	7000	0.0	0.1	½
July 5	308	4400	6500	0.9	0.6	½
July 6 until 1000	328	5000	—	0.8	0.4	½
Average for undisturbed days	298	5820	8250	0.1	0.2	0

[1] American magnetic character figure, based on observations of seven observatories.

[2] An estimate of the severity of the ionosphere storm at Washington on an arbitrary scale 0, ½, 1, 1½, and 2, the character 2 representing the most severe disturbance.

Table II shows the number of hours the night $f_F{}^x$ and daytime $f_{F_2}{}^x$ differed from the average for the undisturbed days of July by more than the given percentage. The disturbed hours are those listed in Table II.

TABLE II

CRITICAL FREQUENCY VARIATION FOR 613 HOURS OF OBSERVATION

Per cent	−60	−50	−40	−30	−20	−10	−0	+0	+10	+20	+30
Number of hours	1	3	6	19	97	200	394	219	34	3	1
Disturbed hours	1	3	6	19	96	164	190	12	6	3	1
Undisturbed hours	0	0	0	0	1	36	204	207	28	0	0

Sudden disturbances of the ionosphere as evidenced by radio fade-outs observed at Washington during July are listed in Table III.

The days during which strong sporadic-E reflections were most prevalent at Washington are listed in Table IV. The table shows the hours of these days during which strong sporadic-E reflections were observed at the given frequencies. The observations were nearly continuous at 4.5, 6, 8, and 10 megacycles. When the frequency is reported as 10 megacycles, this value may have been considerably exceeded.

TABLE III

SUDDEN IONOSPHERE DISTURBANCES

Date	G.M.T.			Location of transmitter	Minimum[1] relative intensity
	Beginning of fade-out	Beginning of recovery	Recovery complete		
July 1	1604	1617	1635	Ontario	0.02
July 1	1905	1914	1920	Ontario, Mass., D. C.	0.0
July 5	1835	1849	1900	Ontario, Mass.	0.0
July 9	1358	1408	1435	Ontario, Mass.	0.0
July 9	1632	1640	1715	Ontario, Mass.	0.0
July 10	1525	1546	1600	Ontario, Mass.	0.0
July 10	1748	1800	1820	Ontario, Mass.	0.0
July 11	1322	1340	1350	Ontario, Mass.	0.1
July 11	1535	1600	1630	Ontario, Mass.	0.0
July 11	1805	1812	1830	Ontario, Mass.	0.0
July 12	1640	1655	1800	Ontario, Mass., D. C.	0.01
July 12	1832	1850	1930	Ontario, Mass., D. C.	0.0
July 13	1448	1500	1525	Ontario, Mass.	0.0
July 19	1942	1948	2000	Ontario	0.02
July 20	1650	1700	1720	Ontario, Mass., D. C.	0.0
July 21	1401	1408	1412	Ontario, Mass., D. C.	0.01
July 23	1618	1624	1640	Ontario, Mass., D. C.	0.01
July 23	1700	1710	1800	Ontario, Mass.	0.05
July 25	1800	1818	1900	Ontario, Mass., D. C.	0.0
July 27	1754	1820	1900	Ontario, Mass.	0.0
July 29	1615	1640	1705	Ontario, Mass., D. C.	0.0
July 31	1718	1726	1740	Ontario, Mass.	0.0
July 31	1746	1800	1900	Ontario, Mass., D. C.	0.0

[1] Minimum relative intensities are in terms of received intensities from CFRX, 6070 kilocycles, 600 kilometers distant.

TABLE IV

SPORADIC E. APPROXIMATE UPPER LIMIT OF FREQUENCY OF THE STRONGER SPORADIC-E REFLECTIONS AT VERTICAL INCIDENCE

Midnight to noon

Date	Hour, E.S.T.											
	00	01	02	03	04	05	06	07	08	09	10	11
July												
6									10	10	8	8
*8	10	10									10	10
12	4.5	4.5	4.5	4.5	4.5	4.5		6	8	10	10	10
13	4.5	6	8	8		4.5	4.5	6	6	8	8	4.5
17	10	10	10	10	10	10	8	4.5	6	4.5	6	4.5
18	10	6	4.5			4.5	4.5	4.5	6	8	4.5	4.5
19	6	6	8	8	8	6	6	6	6	10	10	6
23	6	4.5				8	8	8	8	8	10	8
24	10	8	8	8	4.5		8	4.5	4.5	8	8	6
26	4.5	4.5					6	6	4.5	10	4.5	4.5

Noon to midnight

Date	Hour, E.S.T.											
	12	13	14	15	16	17	18	19	20	21	22	23
July												
3	4.5	4.5				8	8	8	8			
11	4.5		4.5	8	x	x	8	8	8	8	6	6
12	6	4.5	4.5	x		4.5	8	6	8	8	8	6
16					6	4.5		8	10	8	4.5	10
17	4.5	8	4.5	4.5	4.5	4.5	4.5	8	4.5	8	10	10
19	8	8	6	x	4.5	4.5	8	8	8	10	10	10
23	4.5	4.5	4.5	4.5	10		4.5	8	8		8	4.5
24	8	8			8		4.5	4.5	10	10	4.5	4.5
25		4.5	8	4.5	4.5	4.5	10	10	10	10	8	4.5

* No data were obtained at 4.5, 6, and 8 megacycles from July 7 to 10 inclusive.
x No data.

BOOK REVIEWS

Radio-Frequency Electrical Measurements (Second Edition), by Hugh A. Brown. Published by the McGraw-Hill Book Co., 330 West 42nd St., New York, N. Y. 380 plus 5 pages. Price $4.00.

In the seven years which have passed since the appearance of the first edition of this work, new methods of measurement have been devised and new apparatus has rendered possible the refinement of older methods. This second edition is a thorough revision of the book in the light of these facts. Much new material has been incorporated, considerable portions have been rewritten, and the order has been slightly changed. Since this revision has been accomplished without any material increase of the size of the book, it is evident that some of the material of the first edition has been omitted. This has been accomplished without the sacrifice of anything essential.

The general plan of the first edition has been retained. Each section is introduced with a treatment of the underlying theory of the method to be used. Then follow directions for the manipulation, a discussion of sources of possible error, and precautions for their avoidance. The form of presentation of each experiment is based on experience gained by the author with his laboratory classes, and he has endeavored to give sufficient details of procedure to make additional instruction sheets unnecessary for the work of a laboratory course.

The scope of the measurements treated in the book may be indicated by a citation of the chapter headings. These are: Measurement of Circuit Constants; Measurement of Frequency; Antenna Measurements; Electromagnetic Wave Measurements; Measurement of Electron-Tube Coefficients and Amplifier Performance; Modulation, Receiver, and Piezoelectric Crystal Measurements. Primarily the work deals with radio frequencies, but the line is not sharply drawn and standard audio-frequency measuring methods are introduced where their use is appropriate.

The book has been written especially for use in college communications laboratory courses, and the wide scope of the treatment should find for it a usefulness in laboratories quite variously equipped. The book should also be found useful as a manual and guide for the engineer and others who have quantitative measurements to make.

*Frederick W. Grover

* Union College, Schenectady, New York.

The Collected Papers of George Ashley Campbell. Published by the American Telephone and Telegraph Company, New York, N. Y. 548 pages, 11 pages of index.

This volume, privately published as a mark of appreciation on the occasion of his retirement from active service, satisfactorily celebrates Dr. Campbell's distinguished contributions to the science and art of communication. As long as wire communication endures, his name will be remembered as one of the pioneers in developing and applying quantitative mathematical methods to the problems of long-distance telegraphy and telephony. He has lived to see his pioneer achievements become important essential tools of the communication engineer, in daily use in ever-widening fields of application.

While not including all of Dr. Campbell's published work, the most notable omission being the patent specifications on electric-wave filters, nevertheless, all of the developments for which he is principally known are represented in this collection. From the viewpoint of their influence on the communication art, the most important are the 1903 paper on "Loaded Lines in Telephonic Transmission," the 1911 paper on "Cisoidal Oscillations," and the 1923 paper on the "Physical Theory of the Electric Wave-Filter." It is interesting to note that the historical order of development of ideas from "smooth" to "loaded" lines, and thence to selective networks and "filters" has resulted, largely as a result of Campbell's labors, in such powerful and facile methods for the solution of problems in the last-named field that a preferred pedagogical method of presentation at the present time reverses the historical order and proceeds from filters to smooth lines.

With the exception of a group of papers on the perennially controversial question of systems of electrical units, the remaining papers are largely devoted to applications of the ideas and methods of the three papers mentioned above to problems of measurement and the practical operation of equipment.

The volume is provided with a foreword by Vannevar Bush, and an introduction by E. H. Colpitts. Dr. Bush, in addition to a sympathetic appraisal of his achievements, discusses interestingly the position of the research engineer in the modern industrial world as exemplified by Dr. Campbell's career. Mr. Colpitts orients for the reader Dr. Campbell's work in the general development program in the American Telephone and Telegraph Company. The book closes with a complete list of the papers, including the facts as to their multiple publication where such occurred, and an adequate index.

*L. P. Wheeler

* Federal Communications Commission, Washington, D. C.

Radio Operators' Manual (Second Edition). Published by General Electric Company, Radio Department, Schenectady, N. Y. 181 pp. including 4 pp. index, 11 photographs and 45 diagrams. Price $1.00.

This manual is written for those interested in the operation of radio systems for use by police, fire, and conservation departments and by public utilities. The material is particularly well prepared and includes general information relative to the capabilities, applications, and limitations of this class of equipment. The handbook is for the purpose of assisting (a) those preparing for United States Government commercial radiotelephone and radiotelegraph operators' licenses and (b) prospective station licenses.

†H. A. Chinn

† Columbia Broadcasting System, New York, N. Y.

Music and Sound, by Ll. S. Lloyd. Published by Oxford University Press, London, New York, and Toronto. 171 pages including an appendix of 34 pages. Index of 9 pages. 5×9 inches. Price $3.50.

This book is intended for the use of students of music but it may be of interest to others who wish to know why the art of music can use an imperfect scale or the reason for the difference in quality of various musical instruments, In general, books on acoustics are scientific and technical. The treatment in this book is different. The musical scale is explained as it was developed by the composers through the ages. The facts which make keyboard instruments possible, with twelve notes to the octave, are considered in detail. The first portion of the book, concerned with the theory of scales, temperament, consonance, and dissonance should be useful to the engineer who is particularly interested in music from the musician's point of view. The second portion of the book, concerning the quality of musical notes, resonance, stationary waves, the organ pipe, wind instruments, and vibration of strings would be more interesting and useful to engineers if a technical description were employed.

The subjects considered in this book include the pure scale; temperaments; chromatic notes; combination tones; consonance and dissonance; the quality of musical notes; some properties of sound, audibility, sensitiveness of the ear; resonance, forced vibrations, resonators; stationary waves; the organ pipe; orchestral wind instruments; and vibrations of strings.

*Harry F. Olson

* RCA Manufacturing Company, Inc., Camden, N. J.

Magnetron Oscillations of Ultra-Short Wavelengths and Electron Oscillations in General, by Kinjiro Okabe. Published by Shokendo, Japan. 57 pages. Price $1.25.

The object of this little book, says Professor Okabe, is to answer the questions "(1) to what extent have magnetron oscillations of ultra-short wavelengths been developed, and (2) what is the true mechanism of these oscillations."

Five distinct types of magnetron oscillations are defined. Two of the most common types, the so-called "type A" (electron oscillations) and "type B" (sometimes called "resonance oscillations") are described at length by summarizing the work of several investigators. The point is emphasized that the type-B oscillations are fundamentally different from the "negative-resistance" or "dynatron" oscillations obtained at long wavelengths. Single-plate, 2-plate, and 4-plate magnetrons are discussed.

Two chapters are spent on special types of magnetrons. The most important of these is Okabe's "electron-beam magnetron" having split end plates from which the high-frequency energy is taken.

The chapter on mechanism of oscillation gives a good qualitative picture of the mechanism of the types-A and -B oscillations. The theory of the simple "negative-resistance" magnetron is not treated.

The last two chapters are devoted to Barkhausen-Kurz oscillators and to a new type of electron oscillator (Osaka tube) which uses a magnetic focusing field. A complete bibliography on magnetrons is included in the Appendix.

In general the book does not come up to the standard one might expect from a worker who has contributed as much to the development of magnetrons as has Professor Okabe. Possibly due to the extreme brevity and the unusual method of expression many portions of the book lack clarity. The reviewer feels that the book could have been improved by the addition of more quantitative data so as to allow definite comparisons to be made between the various types of magnetrons.

This book will probably be of little value to the general reader but is worth while to the specialist because of the interesting and up-to-date material contained.

*G. Ross Kilgore

* Research and Engineering Department, RCA Manufacturing Company, Harrison, N. J.

The National Physical Laboratory Report for the Year 1937. Department of Scientific and Industrial Research. Published by His Majesty's Stationery Office, Adastral House, Kingsway, London, W.C. 2. Report and appendix 141 pages, index 9 pages, paper cover. Price 2s. 6d. net.

This publication is presented in a smaller form than that of several years ago thereby necessitating considerable reduction in details of the work described. No illustrations or graphs of results are given. The book is divided into the usual nine sections comprising the report of the Executive Committee, which is a summary of the work for the year, with slightly more detailed reports of the different departments of Physics, Electricity, Radio, Metrology, Engineering, Metallurgy, Aerodynamics, and William Froude Laboratory. The papers published in 1937 by individual members of the staff are listed at the end of the report of the department of which the individual is a member. Reference is made in the text to appropriate papers in which the reader may find details of the work.

The report of the Radio Department is given in ten pages. The work is briefly described under four headings: propagation of radio waves, direction finding, measurement of field intensity, and frequency-measuring equipment for the range from 1 to 70 megacycles. Work under the first two headings is further divided into sections on the coefficient of reflection of radio waves at the earth's surface, computation of relative intensities of direct and ionosphere-reflected waves, phase velocity of electromagnetic waves along the ground, design and calibration of Adcock-system direction finders, and a short-wave receiver for a cathode-ray direction finder. The need for further work in field-intensity measurement from 40 kilocycles to 200 megacycles is pointed out after giving the large discrepancies obtained when using seven sets submitted for test.

A portion of the report of the Electricity Department will be of interest to the radio engineer, i.e., the section on electrical standards and measurements. Brief descriptions are given of improvements in the design of standard inductors and quartz-ring oscillators and in the technique of frequency measurement. Measurements on insulating materials, circuit components, and high-frequency currents up to 100 megacycles are mentioned. Other subjects of general interest include brief descriptions of work on the following: static electricity in dry cleaning, magnetic materials, instruments, and high voltage.

There are a number of other subjects in the electrical field, briefly treated in the report of the Physics Department. The remainder of the book is devoted to brief accounts of the work completed or in progress in other branches of science and engineering.

†E. L. Hall

† National Bureau of Standards, Washington, D. C.

Fundamentals of Radio, by Frederick Emmons Terman, Professor of Electrical Engineering, Stanford University, with the collaboration of Lieut. F. W. Macdonald, U.S.N. Published by the McGraw-Hill Book Company, New York. 458 pages. Price $3.75.

This book has been prepared to serve as an introductory radio course. Although it has been written in the same general form and style as "Radio

Engineering" by the same author, this work has approximately only half the pages of the former, and the treatment is simpler with especial emphasis laid on the fundamental principles involved. For its use there is presupposed only an elementary knowledge of alternating-current theory and only simple mathematics is employed in the text. The treatment is, however, sufficiently precise and exacting to fit the book to college use.

The introduction is followed by about fifty pages devoted to circuit elements at high frequencies and a chapter on resonant circuits. The main part of the book is devoted to the fundamental properties of vacuum tubes and their applications in amplifiers, power amplifiers, and oscillators, in the order named. The use of the vacuum tubes as detectors is introduced by a chapter on the theory of modulation.

In chapters illustrative of these general principles are treated sources of power for operating vacuum tubes, radio transmitters, and radio receivers. Considerable attention is paid to the theory of the propagation of radio waves, and this is supplemented by chapters on antennas and radio aids to navigation, and sections on television and acoustics.

Special mention should be made of the unusual number of problems which have been included in each chapter. These have been designed to test thoroughly the student's grasp of the text, but are capable of solution with only a moderate expenditure of time. The book is clearly written and generously provided with figures in the text.

*FREDERICK W. GROVER

* Union College, Schenectady, N. Y.

Television Reception Technique, by Paul D. Tyers. Published by Sir Isaac Pitman and Sons Ltd., London. Obtainable through Pitman Publishing Corporation, New York, N. Y. 144 pages, 85 illustrations. Price $4.00.

To many radio engineers experienced in the sound-broadcast field will come problems of designing apparatus for television reception. For them this book is written. It is the first of its kind and it does the job remarkably well. Written from the engineering angle, it consists of theory and conclusions only that have been proved in actual practical tests. For instance, the best type of antenna, of transmission line, and of reflector for ultra-high-frequency television waves is described.

Designing engineers will appreciate the omission of obsolete circuits and history. The style of writing is clear and concise.

The first chapter deals with basic principles, next aerial systems, signal amplification, then cathode-ray tubes. This is followed by a chapter dealing with the time base (or deflecting circuits) used in television receivers. In this it is believed that the author has overstressed the utility of gaseous tubes used for deflection. Certainly in America the vacuum tube is much more popular than the gas tube for this purpose. We find no reference to the blocking tube oscillator which is widely used in this country.

In other ways the fact that the book is published in Great Britain, and therefore deals exclusively with British television gear does not detract from the usefulness of the text to American engineers, with one exception, and that is the chapter devoted to synchronizing circuits. The type of signal considered is that from the British Broadcasting Corporation transmitter in London. The information, however, is basic, so that it can be applied by the technical reader to equipment designed for transmission according to the Radio Manufacturers

Association standard used by the majority of the experimental transmitters in this country.

The complete design of a receiver is grouped in the chapter "The Vision Receiver." Supplementing the text is a chart which graphically shows the form and the design data of the radio-frequency and intermediate-frequency transformers, deflecting coils, and the like. At the end of the book is a chapter on receiver faults. These are illustrated by numerous photographs of television images showing the most common faults, many of these being labeled "Note also complete lack of interlace," and "There is no interlace." Television engineers realize that good interlacing is a real test of the receiver's ability to synchronize. It would have been desirable, therefore, for the author to include at least one illustration showing a *satisfactory* picture, at least one in which the interlacing was perfect.

This book is the most useful published thus far for that growing group of technical readers who desire authentic information on television receiver design.

*A. F. Murray

* Philco Radio and Television Corporation, Philadelphia, Pa.

Hochfrequenz-Messtechnik, by Otto Zinke. Published by S. Hirzel, Leipzig. 1938. 219 pages plus 3 pages of index, 221 figures in the text. Bound, 15.50 RM; paper 14 RM.

This book, by one of the engineers on the staff of the Institut für Schwingungslehre und Hochfrequenztechnik at the Berlin Technischen Hochschule, is the third in the series, "Physik und Technik der Gegenwart," edited by Professor Fassbender, the director of that Institute.

The book is what we would style a laboratory manual of instruments and methods in the field of high-frequency measurements. The scope of the matters treated is limited by the exclusion of testing methods for transmitters and receivers, of antenna and field-strength-measurement methods, of tube measurement, and of all methods whose use is confined to frequencies in the audible range; it is expected that these will be covered in other volumes in the series.

The material is presented in eight sections, current measurement (34 pp.); voltage measurement (67 pp.); power measurement (4 pp.); frequency measurement (24 pp.); wave-form (harmonic-content) measurement (8 pp.); modulation measurement (25 pp.); measurement of resistance, capacitance, and inductance (42 pp.); and measurements on lines and cables (7 pp.). Each of these sections is prefaced by a brief general discussion of the requirements necessary to insure a given accuracy. Then follow sections on the various instruments and methods in use, arranged (in general) in the order of their precision. The theory underlying the various methods described is not developed in detail, a bibliography of eighty-eight titles being included for reference in this respect.

In general, it may be said that the treatment is representative of modern practice unencumbered by obsolete procedures. As a compact summary of modern methods in the fields covered, the book serves a useful purpose, but in many instances this compactness is carried to an extent which limits its usefulness as a working laboratory tool. As practically all of the instruments and methods discussed are well known in the art in this country, and as adequate discussion of them is available in English, the reviewer can recommend the book to the American engineer only as a luxury and not as a necessity.

†L. P. Wheeler

† Federal Communications Commission, Washington, D. C.

BOOKLETS, CATALOGS, AND PAMPHLETS RECEIVED

The following commercial publications of radio engineering interest have been received by the Institute. You can obtain a copy of any item without charge by addressing the issuing company and mentioning your affiliation with the Institute of Radio Engineers.

CAMERAS · · · *Wholesale Radio Service Company, Inc., 100 Sixth Avenue, New York, N. Y. Catalog 72, 32 pages+cover, 7×10 inches, printed.* This catalog of cameras and photographic supplies supplements the recently issued radio catalog.

CERAMICS · · · *American Lava Corporation, Chattanooga, Tenn. Bulletin No. 36, 8½×11 inches, printed.* Gives electrical and mechanical characteristics of twelve different ceramic materials manufactured by this company and gives dimensions for the shapes most commonly used by the designers of power-type radio-frequency equipment.

ELECTROLYTIC CONDENSERS · · · *Cornell-Dubilier Electric Corporation, Hamilton Boulevard, South Plainfield, N. J. The C-D Condenser, Vol. 2, No. 2, 20 pages, 5½×8⅝ inches, printed.* Contains a brief description and summary of the advantages of etched foil condensers.

LINE-NOISE MEASUREMENT · · · *Tobe Deutschmann Corporation, Canton, Mass. Reprint, 3 pages, 8½×11 inches, printed.* Application of the Tobe noise and fault locator to quantitative measurement of noise introduced into radio receivers by the supply line.

METERS · · · *Roller-Smith Company, 233 Broadway, New York, N. Y. Catalog Supplement 3–48, 2 pages, 8½×11 inches, lithographed.* Additions to the Roller-Smith line of three inch square, flush mounting instruments have recently been made.

OSCILLOGRAPHS · · · *Allen B. Du Mont Laboratories, Inc., 2 Main Avenue, Passaic, N. J. Oscillographer for May and June, 12 pages, 6×9⅛ inches, printed.* This issue is a catalog of Du Mont instruments and cathode ray oscillograph tubes.

RECEIVER MEASUREMENTS · · · *General Radio Company, Cambridge A, Mass. Experimenter for July, 8 pages, 6×9⅛ inches, printed.* Use of the Type 732-A distortion and noise meter for radio receiver testing is discussed. The instrument was originally designed for tests on broadcast transmitters.

RECORD PLAYING EQUIPMENT · · · *Garrard Engineering and Manufacturing Company, Ltd., 17 Warren Street, New York, N. Y. Booklet, 16 pages, 3¾×9⅛ inches, printed.* Lists automatic record changers and phonograph turntable assemblies.

RELAYS · · · *Ward Leonard Electric Company, Mount Vernon, N. Y. Circular 507B, 4 pages, 8½×11 inches, printed.* Descriptions and specifications on a wide line of relays applicable in radio transmitters and vacuum-tube control circuits.

TRANSMITTING EQUIPMENT · · · *E. F. Johnson Company, Waseca, Minn. Catalog 301, 12 pages, 8½×11 inches, lithographed.* Description of antenna coupling units, transmission lines, pressure type condensers and other components for broadcast and commercial radio transmission are described.

Proceedings of the Institute of Radio Engineers
Volume 26, Number 9 *September, 1938*

CONTRIBUTORS TO THIS ISSUE

Benham, Wilfred Earnshaw: Born August 24 1905, at Hove, Sussex, England. Received B.Sc. degree in physics, 1925. Research, University of London, University College, 1925–1926. General Electric Company (England), 1927–1928; assistant valve engineer, International Telephone and Telegraph Laboratories, Inc., 1929–1931; London valve engineer, 1931–1933; engineer in charge of vacuum laboratory, Marconi's Wireless Telegraph Company, 1933–1935; chief of television section, Pye Radio, Ltd., 1935–1936; short-period contracts with Plessey Company and Ultra Electric, 1936–1937; consulting engineer, 1936 to date. Associate member, Institution of Electrical Engineers. Fellow, Royal Society of Arts. Associate member, Institute of Radio Engineers, 1933.

Eckersley, P. P.: Born January 6, 1892, at La Puebla, Mexico. Higher certificate and certificate of technology, Manchester Municipal Technical School. Apprenticed Mather and Platt and Lancashire Dynamo and Motor Company. Wireless equipment officer, Royal Flying Corps, 1915–1918, in Egypt, Salonika, and France. Wireless experimental establishment, R. A. F., 1918. Head of experimental section, design department, Marconi's Wireless Telegraph Company, 1919–1923; chief engineer, British Broadcasting Corporation, 1923–1929. Chairman, Technical Committee Union Internationale de Radiodiffusion, 1924–1935. Vice President, Institute of Radio Engineers, 1937. Member, Institute of Electrical Engineers. Fellow, Institute of Radio Engineers, 1925.

Gilliland, T. R.: See PROCEEDINGS for January, 1938.

Kirby, S. S.: See PROCEEDINGS for January, 1938.

Smith, N.: See PROCEEDINGS for January, 1938.

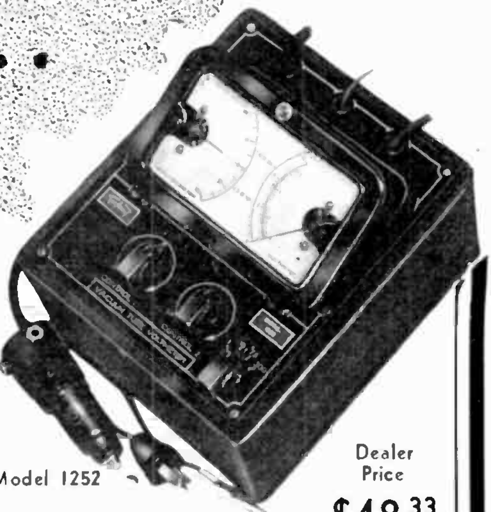
FOR ULTRA

MEASUREMENTS

Use . . .

MODEL 1252 VACUUM TUBE VOLTMETER

Tube On Cable—Particularly Desirable for High Frequency Work . . . Ranges: 3-15-75-300 Volts.



Model 1252

Dealer Price

$48.33

An Outstanding Patented Circuit Recommended by the Best Laboratories and Technicians

This model is self-calibrating, same as other Triplett Vacuum Tube Voltmeters, the self-calibrating feature being automatic with the highly developed tube bridge circuit. It has the exclusive feature for indicating above and below null point. The initial operation of adjusting the bridge at the zero level cancels out error independent of the tube emission values or when replacing tubes.

Model 1252 is furnished with the exclusive Triplett tilting type twin instrument. One instrument indicates when bridge is in balance—other instrument indicates direct reading in peak volts. Furnished complete with all necessary accessories including 1-84, 1-6F5, 1-76. Case is metal with black wrinkle finish. Etched panel is silver and red on black. Dealer Price—$48.33.

Model 1251 same as 1252 but with tube located inside case . . . Dealer Price—$47.67.

Model 1250 same as 1251 except ranges are 2.5, 10, 50 volts. Dealer Price—$36.67.

TRIPLETT
Precision
ELECTRICAL INSTRUMENTS

The Triplett Electrical Instrument Co.
219 Harmon Dr., Bluffton, Ohio

Please send me more information on
......Model 1252; ......Model 1251;
......Model 1250.

Name ....

Address ....

City .... State ....

The Institute of Radio Engineers

Incorporated

330 West 42nd Street, New York, N.Y.

APPLICATION FOR ASSOCIATE MEMBERSHIP

(Application forms for other grades of membership are obtainable from the Institute)

To the Board of Directors

Gentlemen:

I hereby make application for Associate membership in the Institute of Radio Engineers on the basis of my training and professional experience given herewith, and refer to the members named below who are personally familiar with my work.

I certify that the statements made in the record of my training and professional experience are correct, and agree if elected, that I will be governed by the constitution of the Institute as long as I continue a member. Furthermore I agree to promote the objects of the Institute so far as shall be in my power, and if my membership shall be discontinued will return my membership badge.

....
(Sign with pen)

....
(Address for mail)

....
(Date) (City and State)

Sponsors

(Signature of references not required here)

Mr. .... Mr. ....

Address .... Address ....

City and State .... City and State ....

Mr. ....

Address ....

City and State ....

The following extracts from the Constitution govern applications for admission to the Institute in the Associate grade:

ARTICLE II—MEMBERSHIP

Sec. 1: The membership of the Institute shall consist of: * * * (c) Associates, who shall be entitled to all the rights and privileges of the Institute except the right to hold any elective office specified in Article V. * * *

Sec. 4. An Associate shall be not less than twenty-one years of age and shall be a person who is interested in and connected with the study or application of radio science or the radio arts.

ARTICLE III—ADMISSION AND EXPULSIONS

Sec. 2: * * * Applicants shall give references to members of the Institute as follows: * * * for the grade of Associate, to three Fellows, Members, or Associates; * * * Each application for admission * * * shall embody a full record of the general technical education of the applicant and of his professional career.

ARTICLE IV—ENTRANCE FEE AND DUES

Sec. 1: * * * Entrance fee for the Associate grade of membership is $3.00 and annual dues are $6.00.

ENTRANCE FEE SHOULD ACCOMPANY APPLICATION

(Typewriting preferred in filling in this form) No..................

RECORD OF TRAINING AND PROFESSIONAL EXPERIENCE

Name ....
(Give full name, last name first)

Present Occupation ....
(Title and name of concern)

Business Address ....

Permanent Home Address ....

Place of Birth .... Date of Birth .... Age ....

Education ....

Degree ....
(College) (Date received)

TRAINING AND PROFESSIONAL EXPERIENCE

DATES	

Record may be continued on other sheets of this size if space is insufficient.

Receipt Acknowledged .... Elected .... Deferred ....

Grade .... Advised of Election .... This Record Filed ....

POSITIONS OPEN

The following positions of interest to I.R.E. Members have been reported as open on July 30. Make your application in writing and address it to

Box No. ......
THE INSTITUTE OF RADIO ENGINEERS, INC.,
330 West 42nd Street, New York, N.Y.

ENGINEERING SALES

HIGH CALIBRE sales engineer, salary basis, by leading manufacturer of electronic equipment. Box No. 185.

ATTENTION EMPLOYERS . . .

Announcements for "POSITIONS OPEN" are accepted without charge from employers offering salaried employment of engineering grade to I.R.E. members. Please supply complete information and indicate which details should be treated as confidential. Address: "POSITIONS OPEN," Institute of Radio Engineers, 330 West 42nd Street, New York, N.Y.

The Institute reserves the right to refuse any announcement without giving a reason for the refusal.

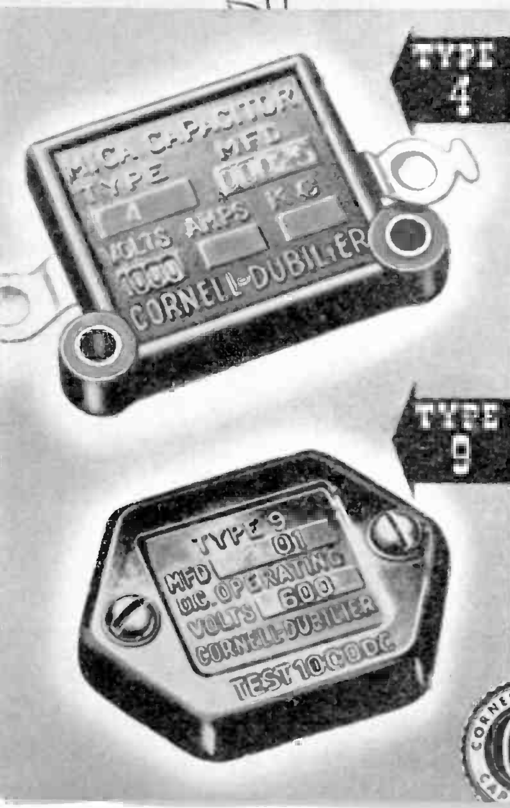
ENGINEERING DIRECTORY

INDEX TO ADVERTISEMENTS

IMITATED • COPIED • BUT NEVER DUPLICATED

C-D *Capacitors* SINCE 1910

THE STANDARD OF QUALITY



THE evolution of the original small "MICADON" capacitor has resulted in the perfection of Types 4 and 9 mica units. Effectively used for r.f. bypass, high voltage D.C. blocking, low power tank capacitors, padders, coupling functions, audio and video purposes. Using the finest quality india ruby mica, select foil and the exclusive C-D stack assembly, these units are hermetically encased in moulded bakelite.

These units are illustrated and described in detail in Catalog No. 160 now ready for distribution.

For consistent dependable results, insist on C-D Capacitors—the pioneer condenser—frequently copied and imitated, but never duplicated.

MICA • DYKANOL • PAPER
WET & DRY ELECTROLYTICS

CORNELL-DUBILIER ELECTRIC CORPORATION
1012 Hamilton Boulevard, South Plainfield, New Jers[ey]
Cable Address: CORDU

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN